



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher or contact*:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk





GenCore version 5.1.6
Copyright. (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 07:50:13 ; Search time 3099 Seconds
(without alignments)
10574 301 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1126

Sequence: 1 gcattattagtggagca.....aataaatgtgtcatgacct 1126

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

1: gb_hg.*
2: gb_hgt.*
3: gb_in.*
4: gb_in.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_to.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_mu.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_to.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rtd.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgt_hum.*
40: em_hgt_mus.*
41: em_hgt_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	853.2	75.8	1752	9	BC002642	BC002642 Homo sapi
2	845.4	75.1	1643	6	AR209918	AR209918 Sequence
3	845.4	75.1	1763	6	AX329664	AX329664 Sequence
4	845.4	75.1	1763	9	HUMCATS	M50696 Homo sapien
5	842.4	74.8	1255	9	HUMCATS	M86553 Human cath
6	842.4	74.8	1255	9	S93414	S93414 cathepsin S
7	669	59.4	1279	10	BC002125	BC002125 Mus muscu
8	667.4	59.3	1305	10	AF038546	AF038546 Mus muscu
9	665	59.1	1296	10	MMCATHEPS	AF002186 Mus muscu
10	631	56.0	1330	10	RATCTHS	L03201 Rattus norv
11	551.4	49.0	1234	4	HICATPS	X62001 H.taurus mk
12	551.4	49.0	1235	4	BOVCATS	M95211 Bovine cath
13	320	28.4	1212	5	CYICYP	L30111 Cyprinus ca
14	313.4	27.8	491	10	MMU18466	Y18466 Mus musculu
15	313.2	27.8	994	4	AF292040	AF292040 Sus scro
16	310	27.5	990	6	AB7768	AB7768 Sequence 2
17	310	27.5	1482	6	AR000416	AR000416 Sequence
18	310	27.5	1482	9	S79895	S79895 OC2-catheps
19	310	27.5	1619	6	AR208697	AR208697 Sequence
20	310	27.5	1619	6	I19038	I19038 Sequence 1
21	310	27.5	1661	9	HS013665	U13665 Human cath
22	310	27.5	1669	6	AR109588	AR109588 Sequence
23	310	27.5	1669	6	AR194312	AR194312 Sequence
24	310	27.5	1669	9	HSOC2RNA	X82153 H.sapiens m
25	310	27.5	1699	9	BC016058	BC016058 Homo sapi
26	308.4	27.4	990	6	AR109592	AR109592 Sequence
27	308.4	27.4	990	6	AR194316	AR194316 Sequence
28	308.4	27.4	990	9	AF070927	AF070927 Macaca fa
29	308.4	27.4	1029	9	AF124092	AF124092 Macaca mu
30	306.8	27.2	990	6	AR109593	AR109593 Sequence
31	306.8	27.2	990	6	AR194317	AR194317 Sequence
32	306.8	27.2	1597	9	HS010280	U20280 Human cath
33	303.8	27.0	1599	4	PABOC2	D14036 Rabbit mRNa
34	300.2	26.7	990	6	AR072914	AR072914 Sequence
35	300.2	26.7	990	6	E28867	E28867 Polynucleot
36	300.2	26.7	990	6	E51957	E51957 Polynucleot
37	300.2	26.7	1248	10	AF010306	AF010306 Rattus no
38	292.4	26.0	990	6	AX305997	AX305997 Sequence
39	292.4	26.0	990	10	MMPPCATHK	X94444 M.musculus
40	289.4	25.7	1044	6	AX063689	AX063689 Sequence
41	262.8	23.3	1420	5	AF358668	AF358668 Oncorhyn
42	256	22.7	1259	3	AF147207	AF147207 Artemia f
43	234.6	21.3	1094	3	AF320565	AF320565 Rhodnius
44	239.2	21.2	1178	9	AF070448	AF070448 Homo sapi
45	239.2	21.2	1342	6	E58936	E58936 Novel human

ALIGNMENTS

RESULT 1
BC002642

LOCUS

DEFINITION

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

BC002642 Homo sapiens, cathepsin S, clone MCC:3886 IMAGE:3610589, mRNA,
complete cds.

BC002642 1 GI-12803614

MCC

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1752)

Strausberg,R.

Direct Submission

Prod. No. is the number of results predicted by chance to have a

JOURNAL

Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc.mgc@nih.gov

Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,

Benjamin, B., Hakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,

Lim, M., Maduro, Q.L., Mastaglio, C., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,

Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,

Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>

Series: IRAL Plate: 12 Row: C Column: 21

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 806607.

FEATURES

source

Location/Qualifiers

1..1752

/organism="Homo sapiens"

/db_xref="locusid:1520"

/db_xref="taxon:9606"

/clone="MGC:3886 IMAGE:3610589"

/tissue_type="Pancreas, adenocarcinoma"

/clone_lib="NIH_MGC_39"

/lab_host="DH10B-K"

/note="Vector: pOTB7"

72..1067

/codon_start=1

/product="cathepsin S"

/protein_id="AAH02642.1"

/db_xref="GI:12803615"

/translation="MKRLVCLVLLVCSVAVALHKDPLDHHHLMKKTYGKQYKENE

EAVRLLEWRNLFVLMHDESHGMSHSDIENHGLDMTSEVMSLSLRVPSOWO

RNIYTKPNWLPDSDVREKGVTEVYKQSCGAWFSAVGALEAGLKDITKLY

SLSQNLYSTKYGKNGCGFMFTAFQYIIDNKGIDSDASYPYKAMDRCQYDSK

YRAATCSKYTELPIYGRDYLKFAVANKPVSVDARHPSPFLYRSYVYEPSCQNV

NHGVLVVGYGDLNCKEYVLVKNWGHNFGEERYIRMARKNKHGFIASPPSPPEI"

545 a 339 c 395 g 473 t

BASE COUNT

ORIGIN

Query Match 75.8%; Score 853.2; DB 9; Length 1752;

Best local Similarity 87.1%; Pred. No. 4.7e-211;

Matches 984; Conservative 0; Mismatches 138; Indels 8; Gaps 4;

QY 1 GCATTATAGTATGGGAGCCTGCAGGTTCTTTC---ATGAATGGCTAGTTGGCTGC 57

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 34 GGACTCTTACGCTGGGAGCACTGCTGGTCTATCAACAATGAACCGGCTGGTTGTGTC 93

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 58 TTCCCTTCTGCTCTATGCTGCACTGGCACAAGTACATAAAGATCCCACTCTGGATCATCACT 117

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 94 TCTTGGTCTGCTCTTCTGAGTGGCAAGTATGATATGAAGATATATCTTGGATCACTACT 153

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 118 GGAATCTCTGGAAGAAACCTACAGCAAAAT---CAAGGAAGAGATGAGGAAGTAGCAC 176

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 154 GGCATCTCTGGAAGAAACCTATGCGCAACATACAGGAAGAAAGATGAAGAAGCAGTAC 213

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 177 GGCCTCTCATCTGGAAAAAATCTAAATTTTGTGATGCTTACATCTGGAACACTCAA 236

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 214 GAGGTCTCATCTGGGAAGAAGATCTAAAGTTTGTGATGCTTCAACACCTGGAGCATCAA 273

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 237 TCGGAATGCATTTCATATCATCTAGGCATGAACCATCTGGGAGACATCACTGGTGAAGAAG 296

Db 274 TGGGAATGCACCTATACGATCTGGGATGAACCACTGGGAGACATGACCACTGAAGAAG 333
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 297 TGATATCTTTGATGGTTCCCTGAGAGTTTCAAGCAATATGAGAGAAATGTCACTATA 456
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 334 TGATCTCTTTGATGAGTTCCCTGAGAGTTCCCAAGCCAGTGGCAGAGAAATATACATA 493
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 357 GGTCAAACTCTAATCAGAAATTCGCTGATCTGCTGACCTGGAGAGAGAGGCTGTGTTA 416
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 394 AGTCAAACTCTAATGATGATATGCTGATCTGCTGCACTGGAGAGAAAGGTTGTGTTA 453
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 417 CTGAAGTGAATACACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 476
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 454 CTGAAGTGAATATCAAGGTTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 513
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 477 TGAAGCACAACTCAAGCTAAAACAGGAAAGTATGCTGCTGCTGCTGCTGCTGCTGCTG 536
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 514 TGAAGCACAACTCAAGCTAAAACAGGAAAGTATGCTGCTGCTGCTGCTGCTGCTGCTG 573
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 537 TAGATTCTCAACTGAAATATGGAATGAAGGCTGCAATGGGCTGCTGCTGCTGCTGCTG 596
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 TGAATTGCTCACTGAAATATGGAATGAAGGCTGCAATGGGCTGCTGCTGCTGCTGCTG 633
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 597 CTTTCCAGTATATTATTAACAACCGGCTTGAATTCAGAAAGCTTCTGCTGCTGCTGCTG 656
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 634 CTTTCCAGTATATTATTAACAACCGGCTTGAATTCAGAAAGCTTCTGCTGCTGCTGCTG 693
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 657 CCATGAATGCAAGTGCAGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 716
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 694 CCATGGATCAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 717 CTGAACCTTCCCTTTGGCAGTCAAGATGCCCTTAAAGAAAGCTGCTGCTGCTGCTGCTG 776
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 754 CTGAACCTTCCCTTTATGGCAGAGAGATGCTGCTGGAAGAGCTGCTGCTGCTGCTGCTG 813
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 777 TGCTCTGGCTATAGATGGCAGCCTATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 836
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 814 TGCTCTGGCTATAGATGGCAGCCTATCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 873
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 837 ATGAACCATCTCTATTCAGAAATGCTGAAATCATGCTGCTGCTGCTGCTGCTGCTGCTG 896
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 874 ATGAACCATCTCTATTCAGAAATGCTGAAATCATGCTGCTGCTGCTGCTGCTGCTGCTG 933
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 897 TTAATGGGAAGACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 956
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 934 TTAATGGGAAGACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 993
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 957 GATATATTCGATGGCAGAAACAGTGGAAATCACTGCTGGGATTCGCTGCTGCTGCTGCTG 1016
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 994 GATATATTCGATGGCAGAAATCACTGCTGGGATTCGCTGCTGCTGCTGCTGCTGCTGCTG 1053
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1017 ACCAGAAATCTA---GACCTCTCTCATTTTATTAACAAGTCCCAAAATTTGAAATGCTTTC 1073
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1054 ACCAGAAATCTAGACGATCTCTCTCTTTTATTAACAATCAAGAAATATCAAGCACTTTC 1113
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATGCTGCTGCTGCTGCTGCT 1122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1114 TCTTAATTTAATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1163
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2

AP209918

LOCUS AP209918 1643 bp DNA linear PAT 20 JUN-2002

DEFINITION Sequence 1 from patent US 6367629.

ACCESSION AR209918

VERSION AR209918.1 GI:21512013

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 1643)

AUTHORS Schneider, P., Yamamoto, K.K. and French, C.K.

TITLE Use of cathepsin S in the diagnosis and treatment of endometriosis


```

|||||
519 CTGAAGTGAATATCAAGGTTCTTGTTGCTTCTGCTGCTTTCAGTCTGTGGGGGCC 578
QY TGAAGCAGCACTGAAGGCTTAAAGCTAGAGAAAGCTGTGTCTGTAGTGCACAGAACTTGG 536
Db TGAAGCAGCACTGAAGGCTTAAAGCTAGAGAAAGCTGTGTCTGTAGTGCACAGAACTTGG 638
QY TAGATTGCTCAACTGAAATATGGAATTAAGGCTGCAATGCGGCTTTCATGACAACTG 596
Db TGAATTGCTCAACTGAAATATGGAATTAAGGCTGCAATGCGGCTTTCATGACAACTG 698
QY CTTTCAGATATATATGATAACAAGGCAATGATTGATGAGAAGCTTTCCTATCCCTACAAG 656
Db CTTTCAGATATATATGATAACAAGGCAATGATTGATGAGAAGCTTTCCTATCCCTACAAG 758
QY CCATGAATGGAAGTGCAGATATGACTCAAAAAGCTGAGCTGCACATGTTCAAGGTATA 716
Db CCATGAATGGAAGTGCAGATATGACTCAAAAAGCTGAGCTGCACATGTTCAAGGTATA 818
QY CTCACCTTCCCTTGGCAGTGAAGATGCTTAAAGAAAGCTGCGCAATAAAGCACCTG 776
Db CTCACCTTCCCTTGGCAGTGAAGATGCTTAAAGAAAGCTGCGCAATAAAGCACCTG 878
QY TGTCTGTGCTATAGATGCGAGCACTATCTTCTTCTTCTGATGAGAAGTGTGTCTACT 836
Db TGTCTGTGCTATAGATGCGAGCACTATCTTCTTCTTCTGATGAGAAGTGTGTCTACT 938
QY ATGAACCATCTCTGATCTCAGATGTAATCATGAGTATGATGCTGCTGCTATGCTAACC 896
Db ATGAACCATCTCTGATCTCAGATGTAATCATGAGTATGATGCTGCTGCTATGCTAACC 998
QY TTAATGGGAAAGCACTACTGCTTGTGAAAGCAAGCTGGGCTCAACTTTGGTGACCAAG 956
Db TTAATGGGAAAGCACTACTGCTTGTGAAAGCAAGCTGGGCTCAACTTTGGTGACCAAG 1058
QY GATATATTCGGATGCGAAGCAAGCTGGAATCATCTGCGGATGCTGCTGCTTCCCTCT 1016
Db GATATATTCGGATGCGAAGCAAGCTGGAATCATCTGCGGATGCTGCTGCTTCCCTCT 1118
QY ACCGAGAAATCTA---GAGCTCTTCATTTTAAACAAAGTCCAAAATATGAAACACTTT 1072
Db ACCGAGAAATCTA---GAGCTCTTCATTTTAAACAAAGTCCAAAATATGAAACACTTT 1178
QY CTCCTTAATTTAATTTTACCTGCTATA-ATAGTAAATAAATATGCTGCTATGA 1122
Db CTCCTTAATTTAATTTTACCTGCTATA-ATAGTAAATAAATATGCTGCTATGA 1229

RESULT 4
LOCUS HUMCATS 1763 bp mRNA linear PRI 19-OCT-2000
DEFINITION Homo sapiens cathepsin S (CTSS) mRNA, complete cds.
ACCESSION M90696.1
VERSION M90696.1 GI:806607
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE
AUTHORS Ritonja, A., Colic, A., Dolenc, I., Ogrinc, T., Podobnik, M. and Turk, V.
TITLE The complete amino acid sequence of bovine cathepsin S and a partial sequence of bovine cathepsin L
JOURNAL FEBS Lett. 283 (2), 329-331 (1991)
MEDLINE 91257334
PUBMED 2044774
REFERENCE
AUTHORS Wiederanders, B., Bromme, D., Kirschke, H., von Figura, K., Schmidt, B. and Peters, C.
TITLE Phylogenetic conservation of cysteine proteinases. Cloning and expression of a cDNA coding for human cathepsin S
JOURNAL J. Biol. Chem. 267 (19), 13708-13713 (1992)
MEDLINE 92317106

```

```

1377692
PUBMED On or before Oct 19, 2000 this sequence version replaced qi:250802,
COMMENT qi:179956.
FEATURES
source Location/Qualifiers
1..1763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="testis"
/tissue_lib="lambda gt11 (Clontech #HL10106)"
1..1763
/gene="CTSS"
137..1132
/gene="CTSS"
/feature="cysteine proteinase; lysosomal enzyme"
/codon_start=1
/product="cathepsin S"
/protein_id="AAC37542.1"
/db_xref="GI:179957"
/translat="MKRLVGLVLCSSAVQLHKDPTLDHRLHLLKKTKYCKYKKNF
FAVRLIWEKLIKFMILHLEHSMGHSYDIGNHLDGDMTSEFVMSLSMLSLRVDSOM
RNITYKSNPNRILPDSVDPEKGVTEKYQSSGACWAFSAVALENJLKLTKGLV
SLSQNLVDSSTEKYNKGNGSPMTTAFQITLONKGLDSASFTKMLKLCYQUSK
YPAATSKYITELPYGREDVLKEAVANKPVSVVDAPHPFFLIYRSGVVEPSTQNV
NHGVLVVCYGDJANGKRYMLVKNMGHNPGEFYIRMARKNKNGHGIASPTYPE1"
137..478
/sig_peptide
479..1129
/mat_peptide
1377692
BASE COUNT 525 a 345 c 400 g 493 t
ORIGIN
Query Match 75.18; Score 845.4; DB 9; Length 1763;
Best Local Similarity 87.28; Pred. No. 5.1e-209;
Matches 986; Conservative 0; Mismatches 136; Indels 9; Gaps 5;
QY 1 GCATTATTAGTATGGGAGCAGCTCCAGGTTCTTTTC---ATGAAATGGCTAGTGGGTGC 57
Db 99 GGACTCTACTGTAGGAGCAACTGCTGTTCTATCACAATGAACGCTGTGTGTGTG 158
QY 58 TTGCTGTGTTGTTCTATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTT 117
Db 159 TCTTGGTGTCTCTCTGCACTGGCAGTTGCAAGAATGCTTACCCTGGATCATCACT 218
QY 118 GGAATTTCTGAGAGAAAGCTACAGTAAAT---AAGAGAGAGAGAGAGAGAGAGAG 176
Db 219 GGTATCTTGGAGAGAAAGCTATGCTAAAT---AAGAGAGAGAGAGAGAGAGAGAG 278
QY 177 GCGCTCATCTGGGAGAGAAAGCTAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 246
Db 279 GAGCTCATCTGGGAGAGAAAGCTAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 348
QY 237 TGGCAATGCAATTCATATGATCTAGTATGAGTATGAGTATGAGTATGAGTATGAG 296
Db 339 TGGCAATGCAATTCATATGATCTAGTATGAGTATGAGTATGAGTATGAGTATGAG 398
QY 297 TGATATCTTGTATGGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 356
Db 399 TGATGTTCTTGTATGGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 458
QY 357 GGTCAAACTCTAAACAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 416
Db 459 AGTCAAACTCTATGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 518
QY 417 CTGAAGTGAATATCAAGGTTCTTGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 476
Db 519 CTGAAGTGAATATCAAGGTTCTTGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 578
QY 477 TGAAGTGAATATCAAGGTTCTTGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 536
Db 579 TGAAGTGAATATCAAGGTTCTTGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
QY 537 TAGATTGCTCAACTGAAATATGGAATTAAGGCTGCAATGCGGCTTTCATGACAACTG 596

```

```

Db      639 TGGATTCCTAAATGAAAAATATGAAACAAAGAGTCTAATGAGTGGCTTATGAAACAG 638
QY      597 CTTTCCAGTATATTATGATAACACCGCATTCATTGATTCAGAGCTTCTTATCCCTACAAAG 656
Db      699 CTTTCATATATATATGATAACAGAGATATGAAATGAAATGAAATGAAATGAAATGAAAT 758
QY      657 CCATGAATGGAATGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 716
Db      759 CCATGGATGCTCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 818
QY      717 CTGAACCTCCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 776
Db      819 CTGAACCTCCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 878
QY      777 TGTCTGTGGGTATAGATGCGAGCAGTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 836
Db      879 TGTCTGTGGGTATAGATGCGAGCAGTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 938
QY      837 ATGAACATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 896
Db      939 ATGAACATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 998
QY      897 TTAATGGAAGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 956
Db      999 TTAATGGAAGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1058
QY      957 GATATATTCGATGCGAAGAACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1016
Db      1059 GATATATTCGATGCGAAGAACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1118
QY      1017 AATAGAAATATATATATATATATATATATATATATATATATATATATATATATATATATAT 1072
Db      1119 ACCCAGAAATCTAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1178
QY      1073 CTTCTAATTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1122
Db      1179 CTTCTAATTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1229

REFERENCE:
AUTHORS      HUMCATSS
LOCUS        Human cathepsin S mRNA, complete cds.
DEFINITION   1255 bp mRNA linear PRI 27-APR-1993
ACCESSION    M86553
VERSION      M86553.1 GI:179958
KEYWORDS     cathepsin; cathepsin S; cysteine protease.
SOURCE       Homo sapiens mature lung cDNA to mRNA.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Cladista; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
              1 (bases 1 to 1255)
              Chapman, H.A.
              Molecular cloning and expression of human alveolar macrophage
              cathepsin S, an elastolytic cysteine protease
              Unpublished (1992)
JOURNAL      Location/Qualifiers
FEATURES     1..1255
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /cell_type="alveolar macrophage cells"
              /tissue_type="lung"
              /dev_stage="mature"
              /germline
              7..1002
              /function="cysteine protease"
              /standard_name="cathepsin S"
              /codon_start=1
              /product="cathepsin"
              /protein_id="AAA35655.1"
              /db_xref="GI:179959"
              /translation="MKRLVGVLLVCSAVAAQLHKDPTLDHIIHLWKTKYQYKKE
              EAVRLIWEKNLAKFVMLNLSHSMHSYDLGNHLDGMTSEEVMSLTSRLVPSQWQ

```

```

BASE COUNT  369 a 253 c 305 g 328 t
OPTIGIN

Query Match 74.8%; Score 842.4; DB 9; Length 1255;
Best Local Similarity 87.5%; Pred. No. 3.2e-208;
Matches 956; Conservative 0; Mismatches 131; Indels 5; Gaps 3;

QY      36 ATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 95
Db      7 ATGAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 66
QY      96 GATCCACCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 154
Db      67 GATCTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 126
QY      155 GAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 214
Db      127 GAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 186
QY      215 CTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 274
Db      187 CTTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 246
QY      275 GGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 334
Db      247 GGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 306
QY      335 TGGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 394
Db      307 TGGCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 366
QY      395 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 454
Db      367 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 426
QY      455 GCTTTCCAGCGCTGCTGGGGCCCTGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 514
Db      427 GCTTTCCAGCGCTGCTGGGGCCCTGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 486
QY      515 TCTCTGAGTGCACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 574
Db      487 ACTCTCAGTGCACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 546
QY      575 AATGGCGCTTTCATGACAACTGCTTTCCAGTATATATATATATATATATATATATATATATATAT 634
Db      547 AATGGCGCTTTCATGACAACTGCTTTCCAGTATATATATATATATATATATATATATATATATAT 606
QY      635 GAAGCTTCTATCCCTACAAAGCCATGATGAAAGTGCAGATATGATGATGATGATGATGATGATGAT 694
Db      607 GAGCGCTTCTATCCCTACAAAGCCATGATGAAAGTGCAGATATGATGATGATGATGATGATGATGAT 666
QY      695 AATGGCGCTTTCATGACAACTGCTTTCCAGTATATATATATATATATATATATATATATATATAT 754
Db      667 GCTGGCAGATGCTCAAGTACACTGAACCTGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 726
QY      755 GCTGTGGCGCAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 814
Db      727 GCTGTGGCGCAATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
QY      815 CTTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 874
Db      787 CTTCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 846
QY      875 TTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 934
Db      847 CTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 906
QY      935 GGCGCTCAACTTGGGTGACCAAGCATATATATATATATATATATATATATATATATATATATAT 994
Db      906 GGCGCTCAACTTGGGTGACCAAGCATATATATATATATATATATATATATATATATATATATATAT

```


ACCESSION VERSION KEYWORDS SOURCE ORGANISM

BC002125
MGC.
house mouse.

REFERENCE AUTHORS TITLE JOURNAL

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1279)
Direct Submission
Submitted (31-JAN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK COMMENT

NIH-MGC Project URL: <http://mgc.nhl.nih.gov>

Contact: MGC help desk

Email: cgaps-remail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILMI at: <http://image.llnl.gov>
Series: IRAK Plate: 8 Row: 4 Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.

FEATURES source

Location/Qualifiers

1..1279

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="CECH 11"

/clone="MGC.6643 IMAGE:3495719"

/tissue_type="Mammary tumor metastasized to lung. Tumor

arose spontaneously from a senescent normal mammary

(clonal) outgrowth infected with the virus MMTV."

/clone_lib="NCI CGAP Lu29"

/lab_host="Df10B"

/note="Vector: pCMV-SPORT6"

24..1046

/codon_start=1

/product="cathepsin S"

/protein_id="AA02125.1"

/db_xref="GI:12805315"

/db_xref="locus:13040"

/translation="MRAPGHAIRWIFMPLVCSVAMHQLQDPTLDYHWDLWKKTTH;
KEYKDNKEEVRLIWEKNKIFIMHNLSEYSGMHYGVGMNDMDIMNEILDKMGA
LPIRQSPKTVTRFSYNSRPLPDTVPKFCVTEVKYQSGCWAFSAVGALEGLG
KLKTKLLSQAQLVDCSNEKYGKNGGAGYMTAFQYIIDNGIHEADASYPKAM
DEKCHNSKRAATCSRYIOLPFGEDALKKAVATKGPVSVGIDASHSSFFFEYKGVY
DIPSCITGVNHCVLVVGYGCTLDGKDYWLKNSGLNFPDQGYIMARNKNHQLASD
CSYPI"

HASH COUNT ORIGIN

374 a 268 c 311 g 326 t

Query Match

Host local Similarity 59.4%; Score 669; DB 10; Length 1279;

Matches 870; Conservative 0; Mismatches 250; Indels 11; Gaps 4;

QY

2 CATTATTAGTATGGAGCAATTCGAGTTCTTTATGAAATGATTAATGGCTGCTCC 61

Db

14 CAGGATGACCATGAGGATCTGAGCAAGTCCCATAGATGGTGTGTTGGATGCCCT 73

QY

62 GTTGCTCCTCATGAGTGGCAGCAAGTACATAAAGATCCCACTCGGATCATCTGAA 121

Db

74 GGTGTGTTTGTGTGTAATGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAG 134

122 TCTCTGGAG 180
134 TCCTGCAAG 193
181 TCCTCATCTGGAG 240
194 TCCTCATCTGGAG 253
241 AATGCAATTCATATGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
254 AATGCAATTCATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 313
301 ATCTTTTAT 340
314 GTGTGGATGAGT 373
361 AAATCTCAATTCAGAAATTCGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
374 ATATCTTAAATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 433
421 AGTGAATACAG 480
434 GGTCAATACAG 493
481 AGTCAATACAG 540
494 AGGCACTGAG 553
541 TTGCTCAACT---GAG 597
554 CTGCTCAATACAG 613
598 TTTCAGAT 657
614 TTTCAGAT 673
658 CATGAATGGAAGTGCAGATATATATATATATATATATATATATATATATATAT 717
674 CATGATGAAG 733
718 TGAATCTCCCTTTGGCAGTGAAGATGCTTTAAAGAGAGAGAGAGAGAGAGAGAG 777
734 TCAGTCTCCCTTTGGTGCAGAGAGATGCTTTAAAGAGAGAGAGAGAGAGAGAG 793
778 GTCTGTGCTATAG 837
794 CTCTGTGGGATGAG 853
838 TGAACCATCTGTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
854 CGACCCCTCTGTACAG 913
898 TAAITGGAG 957
914 TGAATGCAAG 973
958 ATATATTCGAG 1017
974 ATACATTCGAG 1033
1018 CCCAGAAATCTAG 1074
1034 CCCAGAAATCTAG 1089
1075 CTTAATTTAATTTTACCTGCTATATATATATATATATATATATATATATATAT 1125
1090 TCTAATTTAATTTTACCTGCTATATATATATATATATATATATATATATATAT 1140

RESULT 8

AF038546

LOCUS

DEFINITION Mus musculus cathepsin S precursor, mRNA, complete cds.


```

ACCESSION AF038546
VERSION AF038546.1 GI:2746732
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 1305)
AUTHORS Doh-ura, K.
TITLE Mouse cathepsin S cDNA cloning and sequence analysis
JOURNAL Unpublished
AUTHORS Doh-ura, K.
REFERENCE 2 (bases 1 to 1305)
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1997) Dept. of Neuropathology, Neurological
Institute, Kyushu University, 3-1-1 Maidashi, Fukuoka 812-8582,
Japan
FEATURES
source
1..1305
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/tissue_type="brain"
57..1079
/codon_start=1
/product="cathepsin S precursor"
/protein_id="AAH94925.1"
/db_xref="GI:2746732"
/translation="MKAPCHAAIHWLPMPLVCSVAMEQLQRPDITLDYHDLWKKTHE
KEYKNEEVRRLWEKNLKFIMHINLEYSMGHITYQVGMNDGMDNEELSCRMGA
LRISQSPKTVTFRSNRLPTDVTWREKGCVTVEKYGGSCGACWAFSAVGALEGQL
KLTKLKLISLNAQNLVDCSNEEKYGNKGGGYMTFAFYIIDNGGIBADASYPKAM
DKCHYNSKNNAATCSRYIQLPFGDEDALKEAVATKPVSVGIDASHSSFFPYKSGVY
DPPSCTGNVHGVILVVGVTGLDKDYMLVKNWGLNFGDQGYIRMARNNKHCGLASY
CSYPR"
seq_peptide 57..422
mat_peptide 423..1076
/function="cathepsin S"
/feature="cysteine proteinase"
/note="lysosomal enzyme"
HASH COUNT 374 a 275 c 321 g 335 t
ORIGIN
Query Match 59.3%; Score 667.4; DB 10; Length 1305;
Best local Similarity 76.8%; Pred. No. 16;
Matches 869; Conservative 0; Mismatches 251; Indels 11; Gaps 4;
QY 2 CATTATTAGTATGGAGACCTGCGAGGTTCTTTTCATGAAATGGCTAGTTGGCTCTCC 61
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
47 CAGGATGACCATGAGGGCTCTGGCCAGCGCTGCCATAAGATGGCTGTTTGGATGCCCT 106
QY 62 GTTGTGCTCTATGAGTGGCCACAGTACATAAAGATCCCACTCTCGATCATCTAGTGA 121
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
107 GGTGTGCTCTGAGGCTCTGGCCAGCGCTGCCATAAGATGGCTGTTTGGATGCCCT 166
QY 122 TCTCTGGAAGAAATCTAATGAAATCTGATGATGATGATGATGATGATGATGATGAT 180
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
167 TCTCTGGAAGAAATCTAATGAAATCTGATGATGATGATGATGATGATGATGATGAT 226
QY 181 TCTCATCTGGGAAAAATCTAATGAAATCTGATGATGATGATGATGATGATGATGATGAT 240
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
227 TCTCATCTGGGAAAAATCTAATGAAATCTGATGATGATGATGATGATGATGATGATGAT 286
QY 241 AATGATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
287 AATGATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 346
QY 301 ATCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
347 GTCTGGATGGTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406
QY 361 AAATCTTAATAGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

```

407 ATACTTAATGGACATTGCTGACATCTGAGATCTGAGATCTGAGATCTGAGATCTGAG 466
QY 421 AGTCAATATACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
467 GGTCAATATACAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 526
QY 481 AGCACAATGAAATCTGAAATCTGAAATCTGAAATCTGAAATCTGAAATCTGAAATCT 540
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
527 AGGCAGCTGAAGCTGAAGCTGAAGCTGAAGCTGAAGCTGAAGCTGAAGCTGAAGCTGA 586
QY 541 TTGCTCAACT--GAAATATGCGAATAAAGCTGCAATGCGGTTTATGCAAACTGCG 597
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
587 CTGCTCAATGAAAGAAAGTATGGAATAAAGCTGGAAGCTGGAAGCTGGAAGCTGGA 646
QY 598 TTTCAGTATATATTATGATCAACAGCGCATTTGATTGAGAACCTTCTCTATGCTTCA 657
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
647 TTTCAGTATATATTATGATCAACAGCGCATTTGATTGAGAACCTTCTCTATGCTTCA 706
QY 658 CATGAATGAAATGAGATATGATCAACAGCGCATTTGATTGAGAACCTTCTCTATGCT 717
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
707 CATGATGAAAGTGTCTACTATAACTCAAAATGCGAGCTGCTGCTGCTGCTGCTGCT 766
QY 718 TGAACCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
767 TCAGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
QY 778 GTCTGTGCTATAGATGCGAGCGCATTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCT 837
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
827 CTCTGTGCGGATGCGAGCGCGCATTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCT 886
QY 838 TGAACCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
887 CGACCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 946
QY 898 TAATGCGGAAAGACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
947 TGATGCGGAAAGACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1006
QY 958 ATATATTGCGATGCGGAAAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
1007 ATCATTTGCGATGCGGAAAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1066
QY 1018 CCCGAAATCTAGACC---TCTTCATTTTATACAGCTGCGGAAATGCGGAAATGCG 1074
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
1067 CCCGAAATCTAGACC---TCTTCATTTTATACAGCTGCGGAAATGCGGAAATGCG 1122
QY 1075 CTTAATTTAATTTTACCTGCTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1125
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
1123 TCTAATTTAATTTTACCTGCTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1174
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
RESULT 9
MMCATHEPS Mus musculus mRNA 1296 bp linear Ref 06-NOV-1998
LOCUS Mus musculus mRNA for cathepsin S.
ACCESSION AF002386
VERSION AF002386.1 GI:3850786
KEYWORDS cathepsin S.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 1296)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE Direct Submission
JOURNAL Rommelskirch, W.
REFERENCE 2 (bases 1 to 1296)
AUTHORS Submitted (04-NOV-1997) Rommelskirch W., Institut fuer Biochemie,
JOURNAL Klinikum Universitaet Jena, Nonnenplan 1, D-07740 Jena, GERMANY
FEATURES location/Qualifiers
1..1296
/organism="Mus musculus"

```



```
polysignal /product="cathepsin S"
BASE COUNT 352 a 306 c 330 g 342 t
ORIGIN
Query Match 56.0%; Score 631; DH 10; Length 1330;
Best Local Similarity 75.8%; Pred. No. 2.9e-153;
Matches 848; Conservative 0; Mismatches 260; Indels 11; Gaps 5;
QY 14 GGAGACACCTGCGAGTCTCTTCATGAATGGCTAGTTGGGCTGCTTCGCTTGGTCTCTA 73
Db 1 GGACTCTCTCATCGCCCATACGGAGATGGGTGTTTGGGTCGCCCTGGTGTGCTGTCT 60
QY 74 TGCAGTGGCACAGTACATAAAGATCCCACTCTGGATCATCACTGCAATCTCTCTGAAGAA 133
Db 61 GACAAATGGGCACTACAGAGACA-CTTACTCTGAA-AC-AC-ATGAGTCTCTGGAAGAA 119
QY 134 AACCTACA-GCAACAAATCAAGAGAGAGAAATGAGGAAGTAGGACGGGCTCATCTGGGA 192
Db 120 AACTCGCATCAGAGGAATACAGATCAGAAATGAAGAAGACGTACGACGCTCTCATCTGGGA 179
QY 193 AAAAAATCTAATAATTTGTGATGCTTACAACTCTGNACACTCAATGGGAATGCATTCATA 252
Db 180 AAGAATCTCAAGTTTATATGCTTTCACAACTCTGGAACATCTCTATGGGAATGCATTCATA 239
QY 253 TGATCTAGGCATGAACCATCTGGGACATGACTGGTGAAGAGTGCATATCTTTGATGGG 312
Db 240 CAGTGTGGGATGAAACATGAGGACATGACACCTGAAGAAGTACATAGTATATATGG 299
QY 313 TTCCTGAGAGTTCACAGCAATGGGAGAGAAATGACATTAAGTCAAACTTAATCA 372
Db 300 CTCCTCGAATTTCCACAGCCCTCGAATAGAACTGGCACTCTCAAGTCAAGCTTAATCA 359
QY 373 GAAATGCGCTGATTGTGATGTGAGAGAGAGAGGCTGTGTACTGAAGTCAAAATACCA 432
Db 360 GACATTCGCGGACTCTGCTGAGTGGGAGAAAGGCTGCGTCACTAATGTGAATACCA 419
QY 433 GGGTCTTGTGGTCTGTTGGGCTTTAGCGCTCTGAGCGGCGGCGGCGGAGACAACTGAA 492
Db 420 GGGCTCTTGTGTTGCTGTGAGCTTTCAGCGCGGAGGCGGCGGCTTGAAGGCGAGTGA 479
QY 493 GCTAAAAACAGAAAGCTGCTCTGAGTGCACAGAGCTGGTAGTGTGCTCAAC- 549
Db 480 GCTGAAACGGGAGAGCTGCTCTGAGTGCAGAGCTGCTGAGAGCTGCTGAGTGTGCTCAACGGA 539
QY 550 TGAATAATGGGAATAAAGGCTGCAATGGCGGTTTCATGACAACTGCTTTCCAGTATAT 609
Db 540 AGAAAGTACGGGAATAAAGGCTGCGGGGCTGGCTTCATGACCGAAGCTTTCCACTACAT 599
QY 610 TATTGATAACAGCGGCTTGAATGACAAAGCTTCTTATCCCTACAAAGGCAATGAATGGAAA 669
Db 600 CATGATACGA--GATATGAGTACAGAGCTTCTTATGCTTACAAAGGCAATGATGAA 656
QY 670 GTGCAGATGATGATCAAAAAGCGAGCTGCCACATCTTCAAAAGTATACCTGAACTTCCCTT 729
Db 657 GTGCTCTATGACCCAAAATCGAGCTGCCAGCTCTTCAAGGTACATGAGCTCCCTT 716
QY 730 TGCACTGAAGATGCTTTAAAGAACTGTGAGGCAATAAAGAACTGTGTCTGTGAGCTAT 789
Db 717 CGCGGATGAAGAGCTCTGAAAGAGAGCAGTGGCCACATAAAGGCGCTGTCTGTGGGAT 776
QY 790 ---AGATGGGAGCAGTATCTTCTTCTGCTGACAGAGTGGTGTCTACTATGACCAATC 846
Db 777 CGAGGAGGCGAGTATCTTCTTCTTCTGCTGACAGAGTGGTGTCTACTATGACCAATC 836
QY 847 CTGTACTCAGAAATGTAATCATCGAGTATTTAGTGTGGCTATGTTAACTTAACTGAGAA 906
Db 837 CTGTACCGAGAAATGAAATCATGGTCTTCTGCTGTTGGCTATGGGACTCTTTGATGGGAA 896
QY 907 AGATTAATGCTTGTGAAAAAAGTCTGGGCGCTCAACTTTTGTGTGACCAAGGATATATCG 966
Db 897 AGACTACTGGCTTGTGAAAAAAGTCTGGGCGCTTCACTTTTGTGTGATCAAGGATATATTCG 956
```

11
241 CATGTAAATAATCAGCTGCCACTGTTCAAGGTATATGAACCTTCCCTTCGGCAGCGAA 300
740 GATGCTTTAAAGAAAGCTGTGGCAATAAAGGACCTGTGTCTGTGGCTATAGATGGAGC 799
301 CAGGCTTTAAAGAAAGCGTGGCAATAAAGGACCTGTGTCTGTGGTATAGACGCGAGC 360
800 CACTATTCTTCTTCCTGTACAGAGCTGTGTCTACTATGAACCATCTGTACTCAGAAAT 859
361 CATTCTTCTTCTTCCTGTACAAAGCTGTGTCTACTATGACCCCTCCTGCACCTCAGAAAT 420
860 GAGAAATCAAGCTATAGT 919
421 CTGAACCATGTGTACTCGT 480
920 GTGAATCAAGCTATAGT 979
481 CTGAACCATGTGTACTCGT 540
980 GTGAATCAAGCTATAGT 1036
541 CATTCTTCTTCTTCCTGTACAAAGCTGTGTCTACTATGACCCCTCCTGCACCTCAGAAAT 600
1037 TCATTTTATAACAAATTAAGT 1096
601 TTGTTTATAACAAATTAAGT 660
1097 TAATGTAATAAATAAGT 1125
661 TAACAGTACAAATTAAGT 689

RESULT 12
BOVCATS
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
FEATURES
SOURCE
gene
CDS

1235 bp mRNA linear MAM 02-AUG-1993
Bovine cathepsin S (cats) mRNA, partial cds.
M95211.1 GI:162814
cathepsin S; cysteine proteinase; lysosomal enzyme.
Bos taurus spleen cDNA to mRNA.
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (sites)
Ritonja, A., Colic, A., Dolenc, I., Ogrinc, T., Podobnik, M. and Turk, V.
The complete amino acid sequence of bovine cathepsin S and a
partial sequence of bovine cathepsin L
FEBS Lett. 283 (2), 329-331 (1991)
91257334
2044774
2 (bases 1 to 1235)
Wiederanders, B., Broemme, D., Kirschke, H., Kalkkinen, N., Rinne, A.,
Paquette, T. and Tothman, P.
Primary structure of bovine cathepsin S. Comparison to cathepsins
L, H, B and papain
FEBS Lett. 286 (1-2), 189-192 (1991)
91323515
1864368
location/Qualifiers
1. .1235
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue.type="spleen"
1. .1235
/gene="cats"
1. .591
/partial
/gene="cats"
/codon_start=1
/product="cathepsin S"
/protein_id="AAA30435.1"
/db_xref="GI:162815"

translation="CGSWAFSAVGALEAOKVKTGKLVSLSAQNLVDCSTAKYGNKG
CNGCPWTEAFQYIIDNNGIDSPASYPKAMDKCKOYDKNRAATCSRYIEIHPGSEEA
LKEAVANKGPGVGVGIDASSISSELYKTYGVYDPSCTONVNHGVLVVGYNLDSKDWL
VKNSWGLHFGDSQYIIRARNNGHGIANTPSTPEI"
1. .216
/gene="cats"
/product="cathepsin S"
589. .1235
/gene="cats"
1235
/gene="cats"
BASE COUNT 400 a 214 c 242 g 379 t
ORIGIN
Query Match 49.0%; Score 551.4; DB 4; Length 1235;
Best Local Similarity 88.5%; Pred. No. 1.5e-132;
Matches 610; Conservative 0; Mismatches 76; Indels 3; Gaps 1;
QY 440 TGTGGCTTCTTGGGCTTTTCAGCGCTGTGGGGCCCTGGAAGCACTGAAGCTTAA 499
Db 1 TGTGGCTTCTTGGGCTTTTCAGCTGTGTGGGAGCCCTGGAAGCACTGAAGCTTAA 60
QY 500 ACAGAAAGCTGTGTCTCTGAGTGACAGAACTTGGTAGATTGCTCAACTGAAAAATAT 559
Db 61 ACAGAAAGCTGTGTCTCTGAGTGACAGAACTTGGTAGATTGCTCAACTGAAAAATAT 120
QY 560 GGGATAAAGCTGCAATGGCGTTTCATGACAACTGCTTTCCAGTATATATTGATAAC 619
Db 121 GGGATAAAGCTGCAATGGCGTTTCATGACAGAGCTTTCCAAATATATCATTTGACAC 180
QY 620 AACGCCATTGATTGAGAGCTTCCCTATCCCTACAAAGCCCAAGAGTATATCATGATAT 679
Db 181 AATGGCATCGATTGAGAGCTTCCCTATCCCTACAAAGTATATTCAAAGTATATTCAACTTCCCTTCGGCAGCGAA 739
QY 680 CACTCAAAAAAGCGAGCTGCCACATGTTCAAAGTATATTCAAAGTATATTCAACTTCCCTTCGGCAGCGAA 300
Db 241 GATGTAAAAAATCGAGCTGCCACCTGTTCAGAGTATATTCAACTTCCCTTCGGCAGCGAA 799
QY 740 GATGCTTTAAAGAAAGCTGTGGCCAAATAAGAACTGTGTGTGTGTGTGTGTGTGTGTGT 360
Db 301 GAGGCTTTAAAGAAAGCGTGGCCAAATAAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGT 859
QY 800 CACTATTCTTCTTCTCTCTACAAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
Db 361 CATTCTTCTTCTTCTCTCTACAAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 919
QY 860 GTGAATCAAGCTATAGT 480
Db 421 GTGAACCATGTGTACTCGT 979
QY 920 GTGAACCATGTGTACTCGT 540
Db 481 GTGAACCATGTGTACTCGT 1036
QY 980 GTGAACCATGTGTACTCGT 600
Db 541 AGTGGAAATCACTGT 1096
QY 1037 TCATTTTATAACAAATTAAGT 660
Db 601 TTGTTTATAACAAATTAAGT 689

RESULT 13
CYICYP
LOCUS
DEFINITION
ACCESSION
VERSION
CYICYP
Cyprinus carpio cysteine proteinase gene, complete cds.
L30111
L30111.1 GI:463045

BASE COUNT	124 a	111 c	137 g	119 t	
ORIGIN					
Query Match	27.8%; Score 313.4; DB 10; Length 491;				
Best Local Similarity	78.8%; Pred. No. 1.2e-70;				
Matches	387; Conservative 0; Mismatches 101; Indels 3; Gaps 1;				
QY	440	TGTGGTCTGTTTGGATTTTAAAGATGTTGGGGGCTTGGAGACAACTGAAGCTAAAA	499		
Db	1	TGTGGTCTGCTGGCTTTCAGTCTGAGTGGGGCCCTTGAAGGCGAGCTGAAGCTGAAA	60		
QY	500	ATAGCAAACTATGTTCTGCTGAGTGCACAGAACTTGGTAGATGCTCAACT...GAAAAA	556		
Db	61	ACGGGAAGCTGATATGCTCAGTCTCAGAACTTGGTGGAGCTGCTCAATGAAGAAAG	120		
QY	557	TATGGGAATAAAGCTGCAATCGCGGTTTCATGACAACTGCTTCCAGTATATTTGAT	616		
Db	121	TATGGGAATAAAGCTGTGGAGCGGCTACATGACCGAGCTTCCAGTACATATGAT	180		
QY	617	ACAACGGCATTGATTCAGAGCTTCCCTATCCCTACAAAGCCATGAATGGAAAGTGCAGA	676		
Db	181	AATGGGGCATAGAGCGCAGCTTCCCTATCCCTACAAAGCCAGCGATGAAAAGTGTAC	240		
QY	677	TATGATCAAAAAGGAGCTGCCACATGTTCAAGATATACGAACTTCCCTTTGGCAGT	736		
Db	241	TATACTCAAAAATCGAGCTGCCAGCTTCAAGGTACATTCAGCTCCCGTTTGGTGAC	300		
QY	737	GAGATGCCCTTAAAGAGCTGTGGCCATAAAGGACCTGTCTGTGGCTATAGATCGG	796		
Db	301	GAGGATGCCCTGAAAGAGCAGTGGCCACTAAAGGCGCTGTCTGTGGCATCGACGCC	360		
QY	797	AGCCACTATCTTCTTCTGTACAGAGTGTGTCTACTATGACCACTCTGTACTCAG	856		
Db	361	AGCCATCTCTCTCTCTCTACAAAAGCGGTGTCTATGAGGACCCCTCTGTACGGGC	420		
QY	857	AATGGAATCATGGAGTATTAGTGTGTGCTATGTGTAACCTTAATGGGAAGACTACTGG	916		
Db	421	AATGGAATCATGGTGTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	480		
QY	917	CTTGTGAAAAA 927			
Db	481	CTTGTGAAAAA 491			
RESULT 15					
AF292030					
LOCUS	AF292030 994 bp mRNA linear MAM 17-NOV-2000				
DEFINITION	Sus scrofa cathepsin K precursor (CTSK) mRNA, complete cds.				
ACCESSION	AF292030				
VERSION	AF292030.1 GI:10348285				
KEYWORDS					
SOURCE	Sus scrofa.				
ORGANISM	Sus scrofa				
REFERENCE	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.				
AUTHORS	1. (bases 1 to 994)				
TITLE	Cathepsin K in thyroid epithelial cells: sequence, localization and possible function in extracellular proteolysis of thyroglobulin				
JOURNAL	J. Cell. Sci. 113 Pt 24, 4487-4498 (2000)				
MEDLINE	20534869				
PUBMED	11082042				
REFERENCE	2. (bases 1 to 994)				
AUTHORS	Tepel, C., Herzog, V. and Brix, K.				
TITLE	Direct Submission				
JOURNAL	Submitted (02-AUG-2000) Institute of Cell Biology, University of Bonn, Ulrich-Haberland-Str. 61a, Bonn D-53121, Germany				
FEATURES	Location/Qualifiers				
source	1. .994				
gene	/organism="Sus scrofa"				
	/db_xref="taxon:9823"				
	/tissue_type="thyroid"				
	1. .994				

CDS

/gene="CTSK"

2. .994

/gene="CTSK"

/EC_number="3.4.22.38"

/codon_start=1

/product="cathepsin K precursor"

/protein_id="AAG12340.1"

/db_xref="GI:10048286"

/translation="MWGLKVKLVLLPVMSALYPFILLDTOMELWKKTKYKQYNSKVIDE

ISRLIWEKNIKHLISIHNLASLCVHTYELAMNHLCDMTSEVVVOKMTGLKVPSPSHK

SNDLYIPDWEGRTPDSIDYRKGGYTPVANKCCGSCWAFSSVGLFGGLKKTKTKL

LNUSPQNLDCVSENGCGGGMNFAFYVKNRIGIDSEDAFYVGGDENCMNPTGK

AAKRGREYIPEGNEKALKRAVRGVSVAIDASLISFYVSKGVYDNCNSDLN

HVLAVGYSIQGKKHIIKNSWGENGNKGYILMARKNKACGIANLASFPRM"

BASE COUNT 287 a 216 c 270 g 221 t

ORIGIN

Query Match 27.8%; Score 313.2; DB 4; Length 984;

Best Local Similarity 62.3%; Pred. No. 1.3e-70;

Matches 582; Conservative 0; Mismatches 333; Indels 19; Gaps 5;

QY 105 CTGATCATCACTCGAATCTCTGCAAGAAACCTTACAGCAACAAT-CAAGGAAGACAAT 163

Db 68 CTGGATACCCAGTGGGAGCTATGGAAGAGACCTTACAGGAAGCAGTATAACAGCAAGG 127

QY 164 GAGGAAGTAGCACGCGCTCTCATCTGGGAAAAAATCTAAAAATTTGTGATGCTTCACAAT 223

Db 128 GATCAATCTCTCGCGCTTTAATTTGGGAAAAAACCCTGAAGCATATTTCCATCCATAAT 187

QY 224 CTGGAACACTCAATGGGAATGATCATATGATAGTATAGGATAGAACCTATGAGGAGACATG 283

Db 188 CTTGAGGCGCTCTGTGGGTTTACATGATGATGATGATGATGATGATGATGATGATGATG 247

QY 284 ACTGCTGAAGAGTATATCTTTGATGGTTTCCCTGAGAGTTCACAGC---CAATGGCAG 340

Db 248 ACCAATGAGAGTGGTTCAGAGATGATGATGATGATGATGATGATGATGATGATGATG 307

QY 341 AGAATGTGAC---TTATAGTCAACTCTAATCAGAAATTTGTTGATTTCTGTTGATG 397

Db 308 AGTAATGACACCTTTATATCCAGACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 367

QY 398 AGAGAGAGGCGCTGTGTTACTGAGTGAATATATAGGTTTCTTGTGCTGTTGTTGCT 457

Db 368 CGGAAGAGAGATATGTTACTTCTGTCAGAAACAGGGTCACTGTTGTTGTTGTTGTTG 427

QY 458 TTACGCGCTGTGGGGCGCTTGGAGACACAACTGAAGCTAAACACAGCAAGGCTGTGCT 517

Db 428 TTTAGCTCTGTGGGGCGCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 487

QY 518 CTGAGTGCACAGAACTTGGTAGATTGCTCAACTGAAAAATATGGGAATTAAGGCTGCAAT 577

Db 488 CTGAGTCCCGAGAAACCTGGTGGATTGTTCTGCTCTGA-----GAATGATGGCTGTGGA 538

QY 578 GGCGGTTTCACTGACAACTGCTTTCCAGTATATATTATGATACACAGCAGTATGATTCAGAA 637

Db 539 GGGGCTCATAGACCAATGCTTCCAGTATGTCAGAGAGAACCTGGCATTTGATCTAGAA 598

QY 638 GCTTCCCTATCCCTACAAAGCCATGAATGGAAAGTGCAGATATGACTCAAAAAAGCGAGCT 697

Db 599 GATGCTTACCCATAGCTGGGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 658

QY 698 GCCACATGTTCAAAATATACAGTATGCTTTTGGAGTGAATGATGCTTAAAGAACT 757

Db 659 GCTAAGTGCAGAGGTTACAGAGAGATCCCTGAGGGAATGAAAAAGCCCTGAGAGGGCA 718

QY 758 GTGCGCAATTAAGAGCCGTGCTGTGATATATATATATATATATATATATATATATAT 817

Db 719 GTGCGCGAGTGGAGCTGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 778

QY 818 TACAGAGTGGTGT---CTACTATGAACCAATCCCTGTTACTCAGAAATGTGAATCATGAGTA 874

Db 779 TACAGAAAGGTGTGTTATGAGATGAAACTGCAATAGCGATATCTCAACACGCGGTC 838

[illegible]

Search completed: July 1, 2003, 10:20:28
Job time : 3122 secs

Tue Jul 1 14:40:54 2003

DNA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated GC, where differential expression of GS is indicative of GCA. Also included are modulating (M2) GA by contacting GC with an agent that alters the expression of at least one gene in GS; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from GS, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from GS in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPAC at the University of California, San Diego, and is hereby incorporated by reference into the present patent.

Seq	sequence	1763 BP	525 A	345 C	400 G	493 T	0 other:
Query Match		75.1%	Score 845.4	DB 24	Length 1763		
Host local Similarity		87.2%	Pred. No. 7.3e-239				
			0. Mismatches	136	Incls	9	Gaps 5

1	GCATTATTAGTATGTGGTANGACTGTCGAGGTTCTTTTC--ATGAAATGGCTACTGTGGGTGC	57
QY		
99	GGACTCTTATCTGTAGGAGCAACTGCTGGTTCTTATCACAATGAACGGCTGCTTTGTGTGC	158
DB		
58	TTCGGTTGTGCTCCTATGTCAGTGGCCACAAAGTACATAAAGATGCCACTCTGGATCATCACT	117
QY		
159	TC TTGGTGTGCTCTCTGCACTGGCACAGTGTGCATAAAGATCCTACCTCGATCACCACCT	218
DB		
118	GGAAATCTCTGGAGAAAACCTACAGCAAAACAT--GAAGGAAGACGAATGAGGAAGTAGCAC	177
QY		
219	GGCATCTCTGGAGAAAACCTATGCGAAACAATACAAAGAAAAGAATCGAAGACGATAC	278
DB		
177	GGCCTCTCATCTGGAAAAAAATCTTAAATTTTGATGCTTCACAATCTGGAACACTCAA	236
QY		
279	GAQCTCTCATCTGGGAAAAAATCTTAAAGTTTGTGATGCTTCACAACCTTGGAGCATTCAA	338
DB		
237	TGGGAATGCATTCATATGATCTPAGGCATGAACCATCTGGGAGACATGACTGCTGGAAGAAG	296
QY		
339	TGGGAATGCATTCATATGATCTPAGGCATGAACCATCTGGGAGACATGACTGCTGGAAGAAG	398
DB		
207	TCATATCTTTGATGGGTTGCTCTGAGAGTTCCTCAGAGCAATGGCAGAGAAAATGTCATTATA	356
QY		
399	TGATCTCTTTGATGAGTTCCCTGAGAGTTCCTCAGAGTTCCTCAGAGCAATATATCACATATA	458
DB		
357	GGTCAAACTCTAATCAGAAATTTGCTGATTCCTGTGGACTGGAGAGAGAGGCTGTGTTA	416
QY		
459	AGTCAAAACCCCTAATCCGGAATTTGGCTGATTCCTGTGGACTGGAGAGAGAAAGGGTGTGTTA	518
DB		
417	CTCAGACTCAAAATACCAGGGTTCTTCTGGTGCTTCTGGGCTTTTCAGCGCTGTGGGGGCC	476
QY		
519	CTCAGACTCAAAATACCAGGGTTCTTCTGGTGCTTCTGGGCTTTTCAGTGCTGTGGGGGCC	578
DB		

Tue Jul 1 14:40:54 2003

Qy 1073 CTCCTTAATTAAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA 11229

RESULT 4	
ABQ58199	
1D	ABQ58199 standard; cDNA; 431 BP.
XX	
AC	ABQ58199;
XX	
DT	02-AUG-2002 (first entry)
XX	
DE	Human colon cancer related nucleotide sequence SEQ ID NO:1894.
XX	
KW	Human; colon cancer; cancer; tissue profiling; forensic; mapping;
KW	genetic analysis; diagnostic; antisense therapy; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200229086-A2.
XX	
PD	11-APR-2002.
XX	
PF	02 OCT-2001; 2001WA-US40732.
XX	
PK	02-OCT-2000; 2000US-237271P.
XX	
PA	(FARB) BAYER CORP.
XX	
PI	Burgess C, Aslie JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;
PI	Thiagalingam A, Lewis MR;
XX	
DR	WPI; 2002-426115/45.
XX	
PT	New isolated nucleic acid that is differentially expressed in cancer
PT	tissues useful for determining the presence of colon cancer in a cell
PT	or tissue type, and in antisense therapy -
XX	
PS	Claim 1; Fig 1; 796pp; English.
XX	
CC	ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially
CC	expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins
CC	encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be
CC	used in antisense therapy. An antibody immunoreactive with a polypeptide
CC	encoded by (I) is useful for detecting cancer in a patient sample, and
CC	for detecting the presence or absence of a polynucleotide encoded by a
CC	nucleic acid which hybridises to (I) in a cell. A probe/primer derived
CC	from (I) can be used for determining the presence of a nucleic acid which
CC	hybridises to (I), and for determining the phenotype of cells in a sample
CC	of cells from a patient. (I) is useful for determining the presence of
CC	colon cancer in a cell or tissue type, for determining the presence or
CC	state of other type of cancer, in antisense therapy, to generate
CC	macroarrays on a solid surface, to identify a chromosome on which the
CC	corresponding gene resides, and in tissue profiling, forensics, genetic
CC	analysis, mapping and diagnostic applications. (I) can be used to raise
CC	antibodies, and to screen for peptide analogues and antagonists.
XX	
SQ	Sequence 431 BP; 122 A; 84 C; 119 G; 106 T; 0 other;
	Query Match 32.7%; Score 368.6; DB 24; Length 431;
	Best Local Similarity 91.0%; Pred. No. 1.9e-98;
	Matches 392; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY	175 ACGGCGTCTCATCTGGGAAAAAATCTAAATTTGTGCTTCACAACTCGGAACACTC 234
DB	1 ACGACGTCCTCATCTGGGAAAAAATCTAAAGTTTGTGATGCTTCACAACTCGGACATTC 60
QY	235 AATCGGAATGCATTCATATCATCTACGCATCAACCATCTGGGAGACATGACTGGTGAAGA 294

QY 758 GTGGCCAAATAAGAGACCTGTCTGTGCTATAGAGGAGGAGCCACATCTCTTTCTCTG 817
DB 834 GTGGCCGAGTGGACATGTCTGTGTGCTATGATGCAAGCTGACCTCTCTCCAGTTT 893
QY 818 TACAGAAAGTGTCTACTATGAACCATCCCTG---TACTCAGAAATGCAATCATGAGTA 874
DB 894 TACAGCAAGGTGTGTATATGATGAAAGCTTGAATACGATATCTGAACCATGCGTT 953
QY 875 TTAGTGTGTGCTATGTAACCTTAATGGGAAAGACTACTGGCTTGTGAAAGACAGCTGG 934
DB 954 TTGGCAGTGGGATATGGAATCCAGAAAGCAACAGCACTGGATAATTTAAAGACAGCTGG 1013
QY 935 GGCCTCAACTTTGGTGACCAAGGATATATTCGGATGGCAAGAAACAGTGGAAATCACTGT 994
DB 1014 GGAAGAAATGAGGAAAGAAAGGATATATCTCATGGCTCGAAATAGAACACGCTGT 1073
QY 995 GGGATTGTCTAGTATATCCCTCTTACCAGAAATCT 1028
DB 1074 GGCATTGCCAACCTGGCCAGCTTCCCAAGATGT 1107

RESULT 9

ID ABV23393 standard; cDNA; 2042 BP.

AC ABV23393;

XX 16-SEP-2002 (first entry)

XX Human prostate expression marker cDNA 23384.

XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX WO200160860-A2.

XX 23-AUG-2001.

XX 20 FEB 2001; 2001WO-US05171.

XX 17-FEB-2000; 2000US-183319P.

XX 25 MAY 2000; 2000US-207454P.

XX 09-JUN-2000; 2000US-211314P.

XX 18-JUL-2000; 2000US-219007P.

XX 13-DEC-2000; 2000US-255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX Claim 1; Page 4252-4253; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:

CC (a) assessing whether a patient is afflicted with prostate cancer;

CC (b) monitoring the progression of prostate cancer in a patient;

CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;

CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient,

CC (f) assessing the prostate cell carcinogenic potential of a compound;

CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

XX Sequence 2042 BP; 584 A, 467 C; 485 G, 495 T, 11 other.

Query Match 27.5%; Score 310; PR 23; Length 2042;
Best Local Similarity 62.1%; Pred. No. 8.1e-81;
Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

QY 105 CTGGATCATCTCGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAGAGAAT 163
DB 217 CTGGACACCCACTGGGAGCTATGGAAGAACCCACAGGAGCAATATAACAAGGTG 276
QY 164 GAGAAATAGCAGGCTCTCATCTGGAAGAAATATATAATTTGATGCTTCACAT 223
DB 277 GATGAATCTCTGGGCTTTAATTTGGGAAAAAACCTGAAGATATTTCCATCCATAAC 336
QY 224 CTGGAACACTCAATGGGAATGCCATTCATCATCTAGGCATGAACCATCTGGAGACATG 283
DB 337 CTTGAGGCTTCTCTTGTGTCCATAATATGAAATGATGATGATGATGATGATGATG 396

QY 284 ACTGGTGAAGAGTATCTTTGATGGGTTCCCTGAGAGTTCC---ACCAATGCCAG 340
DB 397 ACCAGTGAAGAGGTGGTTCAAGAGTACTGGAATCAAGTCAAGTCAAGTCAAGTCAAG 456
QY 341 AGAAATGTGAC---TTATAGTCAAACTCTAATCAGAAATGCTGATTTCTGTGGATGG 397
DB 457 AGTAATGACACCTTTATATCCCAATGGGAGGTAGAGCCCGACAGCTCTGTCGACTAT 516

QY 398 AGAGAGAGGCTGTGTACTGAAAGTGAATACAGGGTCTTCTGTGTGCTGTGTGGCT 457
DB 517 CGAAAGAAAGGATATGTTACTCTGTCAAAATCAGGCTGCTGCTGCTGCTGCTGCTGCT 576

QY 458 TTCAGCGCTGTGGGGCCCTGGAGACCAACTGAAGCTAAAAACAGCAAGAGCTGTCTCT 517
DB 577 TTTAGTCTGTGGGTGGCTGGAGGCGCAACTGAAGAGAAACTGCAAACTCTTAAAT 636

QY 518 CTGAGTGCACAGAACTTGTGTAGATTGCTCAACTTAAATAATATGAAATATGAAAT 577
DB 637 CTGACTCCCCAGAACCTAGTGTGCTGCTCA-----GAATGATGGCTGTGGA 687

QY 578 GCGGTTTCATGACAACTGCTTCCAGTATATATTGATACACAGCATTGATGACAA 637
DB 688 GGGGTTTACATGACCAATGCTTTTAAATATGTGAGAGAAACCGGGTATTTGACTCTGAA 747

QY 638 GCTTCTATCCCTACAAAGCCATGAATGGAAGTCCAGATATGACTCAAAAACCGAGCT 697
DB 748 GATGCTACCCATATGTGGGACAGGAGAGTGTATGTACACCCACAGGCAAGGCA 807

QY 698 GCCACATGTTCAAGTATACTGAACTTCCCTTTGGCAGTGAAGATGCTTAAAGAGAGCT 757
DB 808 GCTAAATGCAGAGGTACAGAGAGATCCCGAGGGGAATGAGAAAGCCCTGAGAGGGCA 867

QY 758 GTGGCAATAAAGACCTGTGTGCTATAGATCGAGCCACTATCTTCTTCTCTCTG 817
DB 868 GTGGCCGAGTGGGACCTGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 927

QY 818 TACAGAAATGCTGTCTACTATGAACCATCTCTG---TACTCAGATGTGAATCATGGAGTA 874
DB 928 TACAGAAAGGTGTGTATTTATCATCAAAAGCTGCAATAGCAATAATCAACCATCTGCT 987

QY 875 TTAGTGTGCTGTGTTAACTTAAATGGAAGAAATATATATTTTGTATATAATATGTTGG 934
DB 988 TTGGCAGTGGGATATGGAAATCCAGAGGGAACCAAGCACTGATTAATTTAAAGACAGCTGG 1047

QY 935 GGCCTCAACTTTGGTGACCAAGGATATATTCGGATGGCAAGAAACAGTGGAAATCATGCT 994
DB 1048 GAGAAAGTGTGAGAAAGTATATTTATGATGATGATGATGATGATGATGATGATGATG 1107

QY 995 GGCATTGCTAGTATCCCTCTTACCAGCAAAATCT 1028

Example 1: Page 39-40; 62pp; English.

This nucleotide sequence codes for rCatK, a recombinant human cathepsin K (CatK) (see AAM4778) comprising an N-terminal 6His tag and the entire CatK sequence from amino acids 13-329 but with the native active site Cys-139 residue substituted by a serine residue. It was obtained by mutagenic PCR (see AAV05958-61) of CatK cDNA synthesized from macrophage RNA, and cloning into prokaryotic expression vector pQE-30. Immunoassay methods are provided for monitoring patient samples using antibodies against human CatK, optionally in conjunction with a recombinantly produced fragment of human CatK, to detect or monitor the onset or progress of a degenerative bone disease, especially osteoporosis. Antibodies can be raised against an antigenic CatK epitope, especially rCatK or maltose binding protein fused to rCatK-MBP (see AAM4779). An increase in the binding of CatK antibodies to exogenous CatK, e.g. in a patient serum sample, indicates onset or increase of osteoporotic processes. The method is especially used to diagnose and monitor primary osteoporosis. It is rapid and inexpensive and eliminates the need for measurements of bone mass.

SQ Sequence 1556 BP; 424 A; 345 C; 372 G; 415 T; 0 other;

Query Match 27.4%; Score 308.4; DR 19; Length 1556;

Best local Similarity 62.0%; Pred. No. 2.1e-80;

Matches 579; Conservative 0; Mismatches 336; Indels 19; Gaps 5;

105 CTGGATCATCATCGAATCTCTGGAGAAAGAACTACAGCAACAT-CAAGGAGAGAAAT 163
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 64 CTGGACACCATCTGGAGCTATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 164 GAGGAGAGTACGACGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 223
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 124 GATGAATCTCTCGGCGTTTAAATTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 183
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 224 CTGGAACATCAATGGAGTGCATTCATATGATGATGATGATGATGATGATGATGATGATG 283
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 184 CTGAGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 243
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 284 ACTGCTGAAGAGAGTATATCTTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 340
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 244 ACCATGAAGAGAGTGTCTCAG 303
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 341 AGAATATGCTAC---TTATAGTCAACCTTAATCAGAAATTCCTGCTGCTGCTGCTGCTGCT 397
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 304 AGTAATGACACCTTTATATCTCAGATGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 363
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 398 AGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 457
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 364 CGAAGAGAGAGATGCTTACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 458 TTAAAGAGTGTGAG 517
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 424 TTATAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 483
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 518 CTGAGTGCACAGCACTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 577
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 484 CTGAGTGCCTGAG 534
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 578 GCGGCTTTCATCAGCACTGCTTCCAGTATATTTATGATTAACAGCGGCTTGTATTCAGAA 637
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 535 GCGGCTACATCAGCACTGCTTCCAGTATATTTATGATTAACAGCGGCTTGTATTCAGAA 594
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 638 GCTTCTCTATCTTAAAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 595 GATGCTCTACCCATATCTGGAG 654
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 698 GGCACATCTCAAGATATCACTTCCCTTGGCTGAGATGCTGCTTAAAGAGAGAGAGAGAGAG 757
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 655 GCTAATGAG 714
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 758 GTGGCAATAAAG 817
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111

715 CTGCCCGCAGTGGGACCTGCTCTCTGCTGGCCATTCATGCAAGCTGCTCTCTCTCTCTCT 774
 818 TACAAAGATGATGCT 874
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 775 TAAAGATGATGCT 834
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 875 TTACTGCTGCTGCT 934
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 835 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 894
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 935 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 994
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 895 GGACAAACCTGGGCAACAAAGGATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 954
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 995 GGGATTGCTAGTATCT 1028
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111
 955 GGCATTGCCCAACCTGGCCAGCTTCCGCCAAGATGT 988
 111111 111111 111111 111111 111111 111111 111111 111111 111111 111111

RESULT 15
 AAV32610
 ID AAV32610 standard; cDNA; 990 BP.
 XX
 AC AAV32610;
 XX
 DT 26-OCT-1998 (first entry)
 XX
 DE Mutant human cathepsin K (C139A) cDNA.
 XX
 KW Mutant human cathepsin K; protease; SPA; tyrosine phosphatase;
 KW scintillation proximity binding assay; diabetes; cancer;
 KW 4-phosphono(difluoromethyl) phenylalanine group; osteoporosis; ss.
 OS Homo sapiens.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 CDS 1..990
 FT CDS
 FT misc_feature 1..45 /*tag= a /*product= "Mutant human cathepsin K C139A"
 FT /*tag= b /*note= "Encodes the pre-domain"
 FT misc_feature 46..342 /*tag= c /*note= "Encodes the pre-domain"
 FT mat_peptide 343..987 /*tag= d
 FT mutation 415..417 /*tag= e
 FT /*note= "changed from TGT in wild-type to GCT in mutant"
 FT FT
 PN WO9820024-A1.
 PD 14-MAY-1998.
 XX
 PF 03-NOV-1997; 97WO-CA00824.
 XX
 PF 04-NOV-1996; 96US-0030411.
 XX
 PA (MERI) MERCK FROSST CANADA INC.
 XX
 PI Desmarais S, Friesen R, Zamboni R;
 XX
 XX WPI: 1998-348101/30.
 DR P-PSDB; AAW48939.
 XX
 XX Peptide(s) useful in binding assays for tyrosine phosphatases or
 PT cysteine proteases - contain two or more 4-phosphono(difluoromethyl)
 PT phenylalanine groups to improve binding affinity
 XX
 PS Disclosure; Pages 41-42; 59pp; English.

The present sequence represents a mutated human cathepsin K cDNA. The cDNA encodes a mutant human cathepsin K Cl39A protease. The invention provides a method for use in a scintillation proximity binding assay (SPA) for proteases and phosphatases. The method involves using mutated proteases and phosphatases whereby the catalytic cysteine residue of the enzymes are replaced with a serine or alanine residue to correct the problem of interference in SPA from extraneous oxidising and alkylating agents. The mutation affects the catalytic properties of the enzyme but does not affect their binding properties. The invention claims for new ligands for use in SPA which have increased binding affinity for a tyrosine phosphatase or cathepsin protease. The ligands contain at least two 4-phosphono(difluoromethyl) phenylalanine groups which increase binding affinity of the ligand to its respective enzyme. The assay can be used to determine the ability of new ligands and compound mixtures to competitively bind with an enzyme. The method is claimed to allow a better usage of SPA in the discovery of compounds for the treatment and study of diseases, e.g. diabetes, cancer and osteoporosis.

Sequence 990 BP: 287 A; 211 C; 270 G; 222 T; 0 other:

Query Match	27.2%	Score 306.8	DB 19	Length 990
Best local Similarity	61.9%	Prod. No. 5e-80		
Matches 578	Conservative 0	Mismatches 337	Indels 19	Gaps 5
QY	105	CTGGATCATCACTGGAACTCTCTGGAGAAAACCTACAGCAAAACAT-CAAGGAGAGGAAT	163	
DB	64	CTGGACACCCACTGGAGCTATGGAGAAGACCCACAGGAGCAATATATACAACAAGTG	123	
QY	164	GAGGAAGTAGCACGGCGTCTCATCTGGGAAAAAATCTTAAATTTGTGATGTTCCACAAT	223	
DB	124	GATGAATCTCTCGCGCTTTAATTGGGAAAAAAGCTGAAGTATATTTCCATCCATAAC	183	
QY	224	CTGGAAACACTCAATGGGAATGCATTCATATGATCTAGGCATGAACCATCTGGGAGACATG	283	
DB	184	CTTGAGGCTTCTCTTGCTGCCATACATATGAATGGCTATGAACCACTGGGGACATG	243	
QY	284	ACTGNTGAAGAGTGTATATTTTATATAGTTCCTCTGAGATTCGCG---AGCCAATGGCAG	340	
DB	244	ACCACTTGAAGAGTGTGTTTCAAGAATGACTGGACTCAAAAGTAGCCCTGTCTCATTCGCCG	303	
QY	341	AGAAATGTTCAC---TTATAGTCAAACTCTTAATCAGAAATTCCTGATTCCTGGACTGG	397	
DB	304	AGTAATGRCACCTTTATATCCASAATGGGAGGTAGAGCCCGACACTCTGTGCNCTAT	363	
QY	398	AGAGAGAGAGATTTTGTATCTGAAGTGAATATACAGAGTTCTTTGTGTGTCCTTTGGGCT	457	
DB	364	CGAAAGCAAGGATATGTTTACTCTCTGCAAAATCAGGGTCAGTGTGTGTTCCGCTTGGCT	423	
QY	458	TTCAGCGCTGTGAGAGGCTGAGAGCAACATGAAGCTAAAAACAGGAAGCTGTGTCT	517	
DB	424	TTTAGCTTTGTAGTGTGAGTGTGAGAGAGTGAATCAGAGAAAACATGGCAACTCTTAAT	483	
QY	518	CTGAGTGCAGAACTTGGTAGTTGTCTCACTGAAATATATGGGAATTAAGGCTGCCAAT	577	
DB	484	CTGAGTCCCAAGAACTAGTGGATTGTGTCTGA-----GAATGATGGCTGTGA	534	
QY	578	GCGGTTTTTCATACAACTGCTTCCAGTATATATTGATAACAACAGCAATTTATTCAGAA	637	
DB	535	GGGGCTATATGCCCAATGCCCTTCCAATATGTGCAGAAGAACCGGGGTATTTGACTCTGA	594	
QY	648	ATTTCTATTCTTAATAAAGATATTAATGGAAGTGTAGATATGATTAATAAAGAGTACGT	697	
DB	595	GATGCCATGCCATATGTGTGACAGAGAGAGTTGTATGTATCAACGCAACAGCAAGGCA	654	
QY	698	GGCACATGTTCAACATATACGTAACTTCCCTTTTGGCACTCAACATGCCCTTAAAAAGACT	757	
DB	655	GCTAAATGCAGAGGTACAGAGATATCCCGAGGGGAATGAGAAAGCCCTGAGAGGGCA	714	
QY	758	GTAGCTAATAAAGAGTCTGTCTGTGCTATATATGATCGAGCGACTATTCTTTCTTCCTG	817	
DB	715	GTGGCCGAGTGGGACCTGTCTGTGCGCAATTCATGCAAGCCCTGACCTCTTCCAGTTT	774	

GenCore version 5.1.1.b
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic acid nucleic search, using sw model

Run on: July 1, 2003, 08:53:50 ; Search time 1896 seconds
(without alignments)
9618.206 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1126
Sequence: 1 gcaattatagataggagca.....aataaatgtctcagaccat 1126

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 80774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database

1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hic:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_hic:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pin:*
21:	em_gss_vit:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rtd:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715.2	63.5	946	9	AL513990
2	702.4	62.4	892	9	AL548051
3	692.2	61.5	997	9	AL544870
4	682.2	60.6	903	9	AL551671
5	662.6	58.8	1287	11	BC011104
6	572.2	50.8	839	9	AU140074

7	565	2	50	2	848	12	BT539651
8	562.6	50.0	49	7	1013	14	BQ054160
9	560	49.7	1068	14	BQ710837		
10	559.8	49.7	1068	14	BQ710837		
11	557.6	49.5	795	12	HC541942		
12	542.8	48.2	782	12	BQ548684		
13	538	47.8	729	10	AV716038		
14	535.2	47.5	758	13	BT766075		
15	520.6	46.2	1048	10	BE616367		
16	519.2	46.1	799	10	BE616203		
17	513.4	45.6	649	14	BQ309582		
18	513	45.6	804	9	AU138806		
19	512.8	45.5	642	14	BQ309579		
20	512	45.5	640	14	BQ309581		
21	509	45.2	783	13	BM007393		
22	501.4	44.5	853	10	BE531143		
23	500	44.4	954	12	BF795515		
24	485.4	43.1	646	14	BQ327134		
25	484.4	43.0	729	10	AV683234		
26	478.8	42.5	937	14	BQ920640		
27	478.6	42.5	735	12	BG715701		
28	477.8	42.4	612	12	BF044942		
29	476.6	42.3	978	9	AV225053		
30	475.8	42.3	921	14	BQ323641		
31	463.8	41.2	819	13	BT656302		
32	461	40.9	668	14	BQ309567		
33	459.8	40.8	824	13	BT157145		
34	459.6	40.8	644	14	BM820138		
35	458	40.7	978	10	BM614890		
36	457	40.6	866	13	BT694504		
37	455	40.4	877	12	BF340573		
38	454.4	40.4	888	14	BQ960933		
39	453.2	40.2	572	13	BG938409		
40	451.6	40.1	650	12	BG878075		
41	449.2	39.9	596	13	BM511045		
42	445.6	39.6	796	12	BE879042		
43	445	39.5	965	10	BB611859		
44	442.2	39.3	598	14	BQ309576		
45	434.8	38.6	761	12	BG863994		

ALIGNMENTS

RESULT 1
AL513990
LOCUS AL513990 LTI_NFL006.PL2 Homo sapiens cDNA clone CL0BA0112B04 5
DEFINITION AL513990 LTI_NFL006.PL2 Homo sapiens cDNA clone CL0BA0112B04 5
prime, mRNA sequence.

ACCESSION AL513990
VERSION AL513990.1 GI:12777484
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. 946

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL0BA0112B04"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; site 1. Not 1. 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : filianglifetech.com URL :
 http://fulllength.invitrogen.com 3 others
 BASE COUNT 281 a 206 c 243 g 264 t

Query Match 61.5%; Score 692.2; DR 9; Length 997;
 Best Local Similarity 86.4%; Pred. No. 7.5e-191;
 Matches 784; Conservative 3; Mismatches 116; Indels 4; Gaps 2;

Db	315	TGCAGAGAAATATACATATAAGTCAAAACCCCTAATCGGATATTCCTGATTCCTGTCGAC	374
Qy	395	TGAGAGAGAGAGGGCTGTGTACTGAAGTGAATACACAGGGTCTTGTGGTGTGTGGTGG	454
Db	375	TGGAGAGAGAGAGGGTGTGTACTGAAGTGAATATCAAGGTCTTGTGGTGTGTGGTGG	434
Qy	455	GCTTTCAGGCTGTGGGGCCCTGGAGAGACACATGAAGCTAAACACAGAAAGCTGGTG	514
Db	435	GCTTTCAGTGTGTGGGGCCCTGGAGAGACACATGAAGCTAAACACAGAAAGCTGGTG	494
Qy	515	TCTCTGAGTGCACAGAACTTGGTAGATTGCTCAACTGAAATAATATGGGAATAAAGCTGC	574
Db	495	TCTCTCAAGTGCACAACTGGTGGTAGATTGCTCAACTGAAATAATATGGGAATAAAGCTGC	554
Qy	575	AATGGCGGTTTCATGACAACTGCTTCCAGTATATATGATACACAGCGCATTTGATCA	634
Db	555	AATGGTGGCTTCATGACAAAGGCTTCCAGTATATATGATACACAGCGCATTCGACTCA	614
Qy	635	CAAGCTTCTCTATCCCTACAAAGCCATGAATGGAAAGTGCAGATATGACTCAAAAACCGA	694
Db	615	GACGGTTCCTATCCCTACAAAGCCATGATCAGAAATGTCAATATGACTCAAAATATCGT	674
Qy	695	GCTGCCAGATGTTCAAGATATACAGTTCCTTGGCAGTGAAGATGCTTAAAGAA	754
Db	675	GCTGCCAGATGTTCAAGATATACAGTTCCTTGGCAGTGAAGATGCTTCAAGAA	734
Qy	755	GCTGTGGCCAAATAAAGGACCTGTGTGTGGCTATAGATGCGAGCCGCTATCTTCTTC	814
Db	735	GCTGTGGCCAAATAAAGGACCTGTGTGTGGCTATAGATGCGAGCCGCTATCTTCTTC	794
Qy	815	CTGTACAGAGTGGTGTCTACTATCAACCATCTCTACTCAGAATGTGAATCATGGAGTA	874
Db	795	CTGTACAGAGTGGTGTCTACTATCAACCATCTCTACTCAGAATGTGAATCATGGTGA	854
Qy	875	TTAGTGGTGGCTATGATGATACCTTAATGGGAAGACTA	912
Db	855	CTTGTGGTGGCTATGATGATCTTAATGGGAAGATA	892

RESULT 3
 AL544870
 LOCUS
 DEFINITION AL544870 LTI_NFL006_P12 Homo sapiens cdna clone CS0D1012YD10 5
 prime, mRNA sequence.
 ACCESSION AL544870
 VERSION AL544870.1 GI:12877351
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 997)
 Li W.B., Gruber C., Jessee J., and Polayes D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 Bp 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 1..997
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0D1012YD10"
 /clone_lib="LTI_NFL006_P12"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of invitrogen 9800 Medical Center Drive

FEATURES
 source


```

QY 693 GAGTGGACACAT--GTCAAAGTATATGTAACCTTCCCTTTGGCAGTGAAGATG-CCTTAA 749
DB 662 GTCCTGGACATTTCTCAAGAGTACACTGAACCTTCCCTTATGGCAGAGAAGATGCTCGAA 721
QY 750 AAGAGCTGTGGCCCAATTAAGAGGACCTGTGTCTGTGGCTATAGATGCCAG-CCACTATTCT 808
DB 722 AGAAGCTGTGGCCCAATTAAGAGGACCTGTGTCTGTGGCTATAGATGCCAGTTCATCTCTCT 781
QY 809 TTTCTCTCT--GTACAGAAGTGGTGTCTTAC-TATGAACCATCTCTGTACTCA 855
DB 782 TACTTCTCTTACAGAAGTGGTGTCTTACATATGAACCATCTGTACTCA 831

RESULT 8
LOCUS BQ053160 1013 bp mRNA linear EST 29-MAR-2002
DEFINITION AGENCOURT_6821803 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:5934760
5', mRNA sequence.
ACCESSION BQ053160
VERSION BQ053160.1 GI:19812500
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1013)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2121 row: g column: 17
High quality sequence stop: 718.
Location/Qualifiers
1..1013
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5934760"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: blood; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH-MGC Library."
HASH COUNT 295 a 201 c 246 g 271 t
ORIGIN

```

FEATURES

source

```

RESULT 9
LOCUS BQ710837 934 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8495105 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301899
5', mRNA sequence.
ACCESSION BQ710837
VERSION BQ710837.1 GI:21849736
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 934)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D
Email: cgapbs-remail.nih.gov
Tissue procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2518 row: i column: 04
High quality sequence start: 15
High quality sequence stop: 568.
Location/Qualifiers

```

FEATURES

source

```

source
1. .934
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6301899"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI. cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASF_COUNT 261 a 202 c 250 g 221 t
ORIGIN

Query Match 49.7%; Score 560; DB 14; Length 934;
Best Local Similarity 84.4%; Pred. No. 3.1e-152;
Matches 665; Conservative 0; Mismatches 120; Indels 3; Gaps 3;

QY 36 ATGAAATCGCTACTTGGCTTCGTTCCGTTGCTCTCTATCGATGGGCACAAATACATAAA 95
DB 36 ATGAAACGGCTGGTTGTGTGCTTGTGGTGTCTCTCTCGAGTGGGCACAAATGCAATAA 124
QY 96 GATCCCACTCTGGATCATCACTAGCTGGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAG 154
DB 125 GATCCTACCCCTGGATCAACCACTGGCATCTCTGGAAGAAACCTATGCAACAATAACAAG 184
QY 155 GAAGAGAAATGAGGAATAGCAGCGCTCTCATCTGGGAAAAAATCTTAAATTTGTGATG 214
DB 185 GAAAGAATGAGAAGCAGTACGACGCTCTCATCTGGGAAAAAATCTTAAAGTTTGTGATG 244
QY 215 CTTTCAAACTTGGAAACCACTCAATGGGAATGCATTCATATGATCTAGGCATGAACCATCTG 274
DB 245 CTTTCAAACTTGGAGCATTCATGGGAATGCACCTCATACGATCTGGGCATGACCACTG 304
QY 275 GGAGACATGACTGGTGAAGAAAGTGATATCTTTGATGGGTTCCCTGAGAGTTCGACGCCAA 334
DB 305 GGAGACATGACCACTGAAGAAAGTGATGCTTTTATGATGTTCCCTGAGAGTTCGACGCCAC 364
QY 335 TGGTACAGAAATGTCTAATTATAGCTAAACTCTAAATCAGAAATTCCTGATTCCTGTGCAC 394
DB 365 TGCAGAGAAATATCATATATAGTCAAACTTAATGGATATTCCTGATTCCTGTGCAC 424
QY 395 TGCAGACAGAAAGGCTGTGTTACTGAAGTGAATACCAAGGTTCTTGTGCTGTTGTTGG 454
DB 425 TGAGAGAGAAAGGGTGTTACTGAAGTGAATATATCAAGGTTCTTGTGCTGTTGCTGG 484
QY 455 GTTTTATGCTTGTGAGAGGCTTGAAGTCAAACTGAAGCTTAAATCAGAAAGCTGGTG 514
DB 485 GCTTTTCACTGCTGGGGGCGCTGGAAGCACAGCTGAAGCTGAAACAGAAAGCTGGTG 544
QY 515 TCTCTGAGTGCAGAAACTTGTGTAGATTGCTCAACTGAAAAATATGGGAATAAAGGCTGC 574
DB 545 TCTCTAGTGCAGAAACTTGTGTAGATTGCTCAACTGAAAAATATGGGAATAAAGGCTGC 604
QY 575 AATGGCGCTTTCATGACAAGCTGTTTCCAGTATATTATTCATAACAA- CGGCATTGATTC 633
DB 605 AATGGTGGCTTCATGACAAAGGCTTTTCCAGTATATTATTCATAACAAAGGCGCATCGACT 664
QY 634 AGAAGCTTCTATCCTTACAAAGCATGAAATGAAAGTGPAGATGACATCAAAAAAGCG 693
DB 665 AGAGCGTTCTATCCCTTACAAAGCCATGGAATGCAATGCGCAATATGACATCCCAATATCG 724
QY 694 AGTGTGCACATGTTCAAAATATAT-CAACTTCTCTTTGG-ATGTGAAGTGCCTTAAAG 752
DB 725 GGTGTGCATATTTTAAATATATGAAATCTTTATAGGAAATAAAGATGTCCCGAAG 784
QY 753 AAGCTGTGGCCAATAAAGGACCTGTGTCTGTGGCTATPAGATGCGAGCCACTATCTTTCT 812
DB 785 AAGCTGTGGCCAATAAAGGCGCGTGTGTCTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 844

```



```

LOCUS      BG548684                782 bp    mRNA    linear    EST 04-APR-2001
DEFINITION 602576491F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4704407 5',
mRNA sequence.
ACCESSION  BG548684
VERSION     BG548684.1 GI:13547349
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 782)
NIH-MGC http://mgc.nhl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rcp@bbs.ram.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI544 row: f column: 24
High quality sequence stop: 742.
FEATURES             Location/Qualifiers
     source           1..782
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4704407"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B (TI phage-resistant)"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor sequence:
                     5'-ATCTAGAGCGCGGCGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      228 a 166 c 193 g 195 t
ORIGIN
Query Match      48.2%; Score 542.8; DH 12, Length 782,
Best local Similarity 86.2%; Pred. No. 30-147;
Matches 636; Conservative 97; Mismatches 97; Indels 5; Gaps 3;

QY 1 GCATTATTAGTATGGAGCACTGTCAGGTCTTTC--ATGAAATGGCTAGTTGGCTGC 57
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 46 GGACTCTTACTGTAGGAGCACTGCTGGTTCATACATGAAACGGCTGTTGTGTGC 105
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 TTCGTTGTCTCTATGTCAGTGGCAGTACATAAAGATCCCACTCTGGATCATCACT 117
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 106 TCTTGTGTGTG-TAT-TTCATGTCACAGTTCGATAGATCCTACCTGGATCACCCT 165
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 GCAATCTCGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGATGAGGAAGTAGAC 176
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 166 GGCATCTCGGAAGAAACCTATGGCAACAATACAGGAAGAAATGAAGAGAGTAGAC 225
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 177 GCGTCTCATCTGGGAAAAAATCTAATATTTGTCATCTCACAATCTGGNACTCAA 236
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 226 GAGCTCTCATCTGGGAAGAAGATCTAAGATTTCATGCTTCACACCTGGAGCATCAA 285
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 237 TGGGAATGATTCATATGATCTAGGATGAACATCTGGGATACATCTGTGAAGAG 296
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 286 TGGGAATGACATCATAGATCTGGGATGAACACCTGGGAGACATGACCATGAAGAG 345
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 297 TCATATCTTTGATGGTTCCTTCAGAGTTCACGCAATGGCAGAGAAATGTCATTATA 356
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 346 TGATGCTTTTGTAGTTCCTTCAGAGTTCACGCAATGGCAGAGAAATATCATATATA 405

```

```

QY 357 GGICAAACTCTAATCAAAAATTGCTGATTTCTGTGAC-IGGAGAGAGAGAGAGTGTGTTA 416
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 406 AGTCAAAAGCTTAATCGGATATTTGCTGATTCGTGGACTGGAGAGAGAAATGGTGTCTTA 465
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 417 CTGAAGTCAAAATACAGGGTCTTCTGCTGCTTCTGGGCTTTCAGCGCTGCTGGGCGCC 476
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 466 CTGAAGTGAATATATCAAGGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 525
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 477 TGGAGGACAACTGAAGCTAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 516
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 526 TGGAGGACAACTGAAGCTAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 585
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 537 TAGATTGCTCAACTGAAATATATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 596
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 586 TGGATTCTCAACTGAAATATATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 645
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 597 CTTTCCAGTATATTATTGATAACAGCGGATGATTGATGAGAGAGAGAGAGAGAGAGAGAG 656
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 646 CTTTCCAGTATATTATTGATAACAGCGGATGATTGATGAGAGAGAGAGAGAGAGAGAGAG 705
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 657 CCATGATGGAAGTGCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 716
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 706 CCATGATGGAAGTGCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 764
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 717 CTGAACCTTCCTTTGGCA 734
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 765 CTGAACCTTCCTTTATGCA 782
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
LOCUS      AV716038                729 bp    mRNA    linear    EST 11 OCT-2000
DEFINITION AV716038 DCH Homo sapiens cDNA clone ICHAU08 5', mRNA sequence.
ACCESSION  AV716038
VERSION     AV716038.1 GI:10797555
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 729)
XU,X., GU,J., LIU,F., QIU,T., ZHAO,M., LI,Y., HUANG,Q., ZHOU,J.,
SONG,H., GU,Y., YANG,Y., GAO,G., XIAO,H., LI,N., QIAN,B., GAO,X.,
CHENG,Z., XU,S., GU,W., TU,Y., JIA,J., FU,G., PEN,S., ZHOU,M., LU
,G., CHENG,Z., and HAN,Z.
Homo sapiens cDNA DCH clones
Unpublished (2000)
Contact: Zequan Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chac.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source           1..729
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DCHAU08"
                     /clone_lib="DCH"
                     /cell_type="dendritic cells"
                     /dev_stage="mature"
                     /lab_host="BM25.8"
                     /note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiI"
BASE COUNT      212 a 142 c 168 g 204 t
ORIGIN
Query Match      47.8%; Score 538; DH 10; Length 729,
Best local Similarity 86.3%; Pred. No. 7.5e-146;
Matches 605; Conservative 0; Mismatches 94; Indels 4; Gaps 1;

QY 425 AATACACAGGGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 484

```


1	AAATATCAAGGTTCTTGTGGTCTGCTGGGCTTTCAGTGCCTGGGGCCCTGGAAGCA	60
485	CAACTCAAGCTAAAAACAGGAAGCTGGTCTCTCTAGTCCACAGAACTTGGTACATTCG	544
61	CAGCTCAAGCTGAAACAGGAAGCTGGTCTCTCTAGTCCACAGAACTTGGTACATTCG	120
545	TCAACTCAAGCTGAAACAGGAAGCTGGTCTCTCTAGTCCACAGAACTTGGTACATTCG	604
121	TCAACTCAAGCTGAAACAGGAAGCTGGTCTCTCTAGTCCACAGAACTTGGTACATTCG	180
605	TATATTATTGATAAACAAGGCGATTGATTCAGAAAGCTTCTATCCCTACAAAGCCATGAAT	664
181	TACATCATTGATAAACAAGGCGCATCGACTCAGAGCGCTTCTATCCCTACAAAGCCATGAAT	240
665	GGAAGCTGAGATATGACTCAAAAAAGCAGCTGCCACATGTTCAAAGTATACTGAACCT	724
241	CAGAAATGTCATATATGACTCAAAAAATGCTGCTGCCACATGTTCAAAGTACACTGAACCT	300
725	CCCTTTGGCAGTGAAGATGCTTTAAAGAAAGCTGTGGCCAAATAAAGGACCTGTGTCGTG	784
301	CCTTATGGCAGACAAGATGTCCTGAAAGAAGCTGTGGCCAAATAAAGGACCTGTGTCGT	360
785	GCTATAGATGCGAGCCACTATTCCTTCTCTCTACAGAAAGTGGTGTCTACTATGAACCA	844
361	GGTGTAGATGCGCGTCACTCCTTCTCTCTACAGAAAGTGGTGTCTACTATGAACCA	420
845	TCTGTACTCAGAAATGTGAATCATGGAGTATTAGTGGTGGCTATGCTAACCTTAATGGG	904
421	TCTGTACTCAGAAATGTGAATCATGGTGTACTTGTGGTGGCTATGGTATCTTAATGGG	480
905	AAAGACTACTGGCTTTGTGAAAAACAGCTGGGGCCCTCAACTTTGGTGACCAGGATATATT	964
481	AAAGAAATACTGGCTTTGTGAAAAACAGCTGGGGCCCTCAACTTTGGTGAAAGGATATATT	540
965	CGGATGCAAGAACAACAGTGGAAATCACTGTGGGATTCGTAGTTATCCCTCTTACCCAGAA	1024
541	CGGATGCAAGAANAATAAGGAAATCATTTGTGGGATTCGTAGTTATCCCTCTTACCCAGAA	600
1025	ATCTA---GACCTCTTCATTTTATAACAAGTCCAAAAAATGAAACACTTCTCTTAATT	1081
601	ATCTAGAGGATCTCTCTTTTATAACANATCAAGAAATATGAAGCACTTCTCTTAACCT	660
1082	TAATTTTACCTGCTATATAGTAAAAAATAATGTGTCATGA	1122
661	AATTTTTCCTGCTCTATCCAGAGAAATAAATGTGTCATGA	701

RESULT 14	
BI766075	
LOCUS	758 bp mRNA linear EST 25-SEP-2001
DEFINITION	60305305f1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:520244 5', mRNA sequence.
ACCESSION	BI766075
VERSION	BI766075.1 GI:15757653
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 758)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: rgaphs-r@mail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

```

Plate: L1A11507 row: f column: 13
High quality sequence stop: 752.
Location/Qualifiers
1. .758
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5202444"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SORT6;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026 Note
this is a NIH_MGC Library."
226 a 161 c 192 g 179 t

```

BASE COUNT	226 a	161 c	192 g	179 t	
ORIGIN					
Query Match	47.5%, Score 535.2, DB 13, Length 758;				
Best Local Similarity	85.6%; Pred. No. 5e-145;				
Matches 630; Conservative	0; Mismatches 103; Indels 3; Gaps 3;				
Qy	36	ATGAATGGCTAGTTGGCGTCTCCGTTGTGCTCTCTATGAGTGGCAACTACATAAA	95		
Db	21	ATGAACGGGTGTGTG-TGCTCTGGTGTGCTCTCTGAGTGGCACAGTTGCATAAA	79		
Qy	96	GATCCCACTCTGGATCATCACTGAATCTCTGGAAGAAACCTACAGCAAACAAT-CAAG	154		
Db	80	GATCTACCTGGATCAACCACTGCGATCTCTGGAAGAAACCTATGGCAACAATACAAG	139		
Qy	155	GAAGAGAAATGAGGAAGTAGACAGCGCTCTCATCTGGGAAAAAAATCTAAAAATTTGTGATG	214		
Db	140	GAAGAAGATGAAGAAG-AGTAGACAGTCTCATCTGGAAGAAATCT-AAAAGTTGTRATG	198		
Qy	215	CTTCACAATCTGGAACACTCAATCGGAATGCATTCATATGCTAGGCATGAACCATCTG	274		
Db	199	CTTCACAACCTGGAGCATTCATGGGAATGCAATCATATGCTAGGCATGAACCATCTG	258		
Qy	275	GGAGACATGACTGGTGAAGAAGTGATATCTTTTGATGGTGTCCCTGAGAGATTCGACGCAA	334		
Db	259	GGAGACATGACAGTGAAGAAGTGATGTCTTTTGATGAGTTCCCTGAGAGATTCGACGCG	318		
Qy	335	TGGCAGAGAAATGTCACTTTATAGGTCAAACTCTAATCAGAAATTCGCTGATTTCTGTGCAC	394		
Db	319	TGGCAGAGAAATATCACATATANGTCAAAACCTAATCGATATTCGCTGATTTCTGTGCAC	378		
Qy	395	TGGAGAGAGAAAGGCTGTGTTACTGAAGTGAATACCAAGGTTCTTTGTGGTCTGTTTGG	454		
Db	379	TGGAGAGAGAAAGGCTGTGTTACTGAAGTGAATATCAAGGTTCTTTGTGGTCTGTTTGG	438		
Qy	455	GCTTTCAGCGCTGGGGGCCCTGGAAGTCAACTGAAGCTAAAAACAGGAAGCTGGTG	514		
Db	439	GCTTTCAGTGTGGGGGCCCTGGAAGCACAGCTGAAGCTGAAAACAGGAAGCTGGTG	498		
Qy	515	TCTCTAGTGGCACGAACCTTGGTAGATTGCTCAACTGAAAAATATCGGAATAAAGCTGC	574		
Db	499	TCTCTAGTGGCCAGAACCTGGTGGATTGCTCAACTGAAAAATATGGAACAAAGGCTGC	558		
Qy	575	AATGGCGGTTTCATGACAACTGCTTTCAGTATATATTGATTAACACCGCATTTGATTCA	634		
Db	559	AATGGTGGCTTCATGACAAAGGCTTTCAGTATATTTGATTAACAAAGGCGATTCGACTCA	618		
Qy	635	GAAG-TTCTATGCTACAAAATCATGAATACAAATG-AGATATGATCAAAAAATGGA	694		
Db	619	GACGCTTCTATCCCTACAAAGCCATGGATCAGAAATGTCAATATGACTCAAAATATCGT	678		
Qy	695	GCTGCCCATGTTCAAAGTATACTGAACCTTCGCTTTGGCAGTCAAGATGCCCTTAAAGAA	754		
Db	679	GCTGCACATGTTCAAAGTACACCTGAATCTCTTATGAGAGAAATCTTTGAAAGAA	738		

```

QY 755 GCTGTGGCCCAATAAAG 770
Db 739 GGCTGTGGCAATAAAG 754

RESULT 15
BE616367
LOCUS 1048 bp mRNA linear EST 24-AUG-2000
DEFINITION 601279178F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3611275 5',
mRNA sequence.
ACCESSION BE616367
VERSION BE616367.1 GI:9897966
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1048)
NIH-MGC http://mgi.nhl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rqbbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM268 row: k column: 20
High quality sequence start: 18
High quality sequence stop: 705.
Location/Qualifiers
1..1048
/orqanism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3611275"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT 293 a 227 c 296 g 231 t 1 others
ORIGIN

Query Match 46.2%; Score 520.6; DB 10; Length 1048,
Best Local Similarity 84.5%; Pred. NO le-140;
Matches 668; Conservative 0; Mismatches 114; Indels 9; Gaps 7;

QY 3 ATTATTAGTATGGGACCACTGCAGTCTTCTTC---ATCAAAATGGCTAGTGGGCTGCTT 59
Db 48 ACTCTTACCGTGGGAGCAACTGCTGTCTTATACAAATGAACGGCTGTGTGTGCTC 107
QY 60 CGCTTGCTCTATGCTGTCAGTGGCAGGTACATCAAGATCCACATCTGATCATCATCTGG 119
Db 108 TTGGTGTGCTCTCTGCTGTCAGTGGCAGTGTGCATAAAGATCTTACCCTGGATCACCAC 167
QY 120 AATCTCTGGAAGAAACCTTACAGCAACAAAT-CAAGGAAGAGAAATGAGGAAGTAGCAGG 178
Db 168 CATCTCTGGAAGAAACCTTATGGCAACAAATCAAGGAAGAAAGAAATGAAGAGCAGTACGA 227
QY 179 CGTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTTCACAAATCTGGAACACTCAATG 238
Db 228 CGTCTCATCTGGGAAAAAGATCTAAAGTTGTGATGCTTTCACAAACCTGGAGCATTCAATG 287

```

```

QY 239 GGAATGCAATTCATATGATCTAGGACATGAAACATCTGGGAGACATGACTGCTGGAAGAGTG 298
Db 288 GGAATGCAATTCATATGATCTAGGACATGAAACATCTGGGAGACATGACTGCTGGAAGAGTG 347
QY 299 AATATCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 358
Db 348 ATGCTCTTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 407
QY 359 TCAAACTCTTAATCAGAAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418
Db 408 TCAAACTCTTAATCAGAAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 467
QY 419 GAATCTGAAATATTAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 478
Db 468 GAATCTGAAATATTAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 527
QY 479 GAAGCAAACTGAAAGCTAAAAACAGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 537
Db 528 GAAGCAAACTGAAAGCTAAAAACAGGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 587
QY 538 AGATTGCTCAACTCAAAATATGGAATTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596
Db 588 AGATTGCTCAACTCAAAATATGGAATTAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 647
QY 597 GTTCTTCTATATTATTGATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 656
Db 648 GTTCTTCTATATTATTGATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 706
QY 657 CCATCAATGGAAGCTGCGACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 715
Db 707 CCATCAATGGAAGCTGCGACATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 766
QY 716 ACTCAATCTTCTCTT-TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 774
Db 767 ACTCAATCTTCTCTT-TGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
QY 775 TGTCTCTGCTG 785
Db 827 GGTCTGTGCTG 837

```

Search completed: July 1, 2003, 10:52:23
Job time : 1911 secs

us-10-010-580-1.rni

Tue Jul 1 14:40:54 2003

GenCore version 5 1 6

om nucleic - nucleic search, using sw model

Run on: July 1, 2003, 09:15:25 ; Search time 70 Seconds
(without alignments)
4933.115 Million cell updates/sec

Title:	US-10-010-580-1
Perfect score:	1126
Sequence:	1 qcattattagtatggagca.....aataaatgtgtcatgacct 1126

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapexl 1.0

Searched: 441362 seqs, 153338381 residues

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database :
Issued_Patents_NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5H_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6R_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PTN5_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfiles1.seq:*
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Query		ID	Description
	Score	Match Length DB		
1	845.4	75.1	1643	US-09-701-685-1
2	310	27.5	1482	US-08-330-121B-1
3	310	27.5	1482	PCT-US95-13820-1
4	310	27.5	1619	US-08-208-007A-1
5	310	27.5	1619	US-08-915-095A-1
6	310	27.5	1619	US-08-798-096-1
7	310	27.5	1619	US-08-798-095A-1
8	310	27.5	1619	PCT-US94-04781-1
9	310	27.5	1669	US-08-964-308-3
10	310	27.5	1669	US-08-964-313-3
11	310	27.5	1669	US-09-069-13A-3
12	308.4	27.4	990	US-08-964-308-12
13	308.4	27.4	990	US-08-964-313-12
14	308.4	27.4	990	US-09-069-13B-12
15	306.8	27.2	990	US-08-964-308-13
16	306.8	27.2	990	US-08-964-313-13
17	306.8	27.2	990	US-09-069-13B-13
18	306.8	27.2	1614	US-08-684-932A-35
19	300.2	26.7	990	US-08-806-959-1
20	236	21.0	1366	US-08-883-526-2
21	162.2	14.4	651	US-08-546-712-1
22	162.2	14.4	651	US-08-751-105-1
23	137	12.2	1390	US-08-821-994-61
24	137	12.2	1441	US-08-821-994-63
25	133.8	11.9	1434	US-08-821-994-62
26	132.2	11.7	1102	US-08-821-994-86
27	130.2	11.6	1661	US-08-821-994-82

ALIGNMENTS

RESULT 1
US-09-701-685-1

Patent No. 6387629
GENERAL INFORMATION:
APPLICANT: Schneider, Patrick
APPLICANT: Yamamoto, Karen K.
APPLICANT: French, Cynthia K.

Query Match	75.18;	Score 845.4;	DB 4;	Length 1643;
Best Local Similarity	87.28;	Pred. NO. 9.5e-251;		
Matches 986:	Conservative	0;	Mismatches 136;	Indels 9;
	Gaps			5;

QY	1	GCATTAATTASTATGGAGCACCTCAGGTTCTTTTC---ATGAAATGGCTAGATTGGGTGTC	57
Db	32	GGACTCTTACTCTAGGAGCAACTCGTGTTCTATCAACAATGAAACGGCTCGGTTTGTGTGC	91
QY	58	TTCCGTTGTGCTTCCTATCGACTGGCACAAGTACATAAAGATCCACCTCTGGATCATCACT	117
Db	92	TCCTTGTGTGCTTCCTCTCGAGTGGCACAAGTTGCATTAAGATCTCACCCTGGATCACCCT	151
QY	118	GGAATCTCTCGAAGAAAACCTACAGCAACAAT--CAAGGAAGAGAAATGAGGAAGTAGCAC	176
Db	152	GGCATCTCTGGAAGAAAACCTATGCAACAATAACAAGGAAAAGAAATGAAGAAGCAGTAC	211
QY	177	GAGGTCTCATCTCGGNAANAANATCTAANAATTGTATGCTCTTCA--AATCTCGAACTCAA	236
Db	212	GAGGTCTCATCTCGGNAANAAGATCTAAGACTTGTGATGCTTTCACAACCTGGAGCAATCAA	271


```
QY 638 GCTTCATATCCCTACAAAGCCATGAATGGAAAGTGCAGATATCACTCAAAAAAGCGAGCT 697
Db 736 GATCCCTACCATATGTGGACACAGGAAGAGAGTGTATGTACAAACCAACAGCAAGGCA 795
QY 698 GGCACATGTTCAAGATATACATGAACCTCCCTTTGGCAGTGAAGATCGCTTAAAGAGCT 757
Db 796 GCTAATATGAGAGGTGACAGAGATCCCGAGGGGAATGAGAAAGCCCTGAAGAGGCA 855
QY 758 GTGGCCAAATAAGACACCTGTGTCTGTGCTATACATGCGGAGCCACTATTTCTTCTTCCTG 817
Db 856 GTGGCCGAGTGGAGCTGTCTGTGTGGCATTGATGCAAGCCTGACCTCTCTCCAGTTT 915
QY 818 TACAGAAGTGTGTCTACTATGAAACCATCTG---TACTCAGATGTGAATCATGAGTA 874
Db 916 TACAGCAAGGTGTGATTAATGATGAAGCTGCAATAGCCATAATCTGAAACCATGGGTT 975
QY 875 TTACTGTGTGCTATGTGAACCTTAATGGGAAGACTACTGGCTTGTGAAAAACAGCTGG 934
Db 976 TTGCAGTGGATATGAAATCCAGAGGGAACAGCAGCTGGATAATTAATAAACAGCTGG 1035
QY 935 GGCCTCAACTTGTGTGACCAAGGATATATCGGATGCAAGAACAGCTGGAAATCACTGT 994
Db 1036 GGAGAAACTGGGAAACAAAGGATATATCTCTATGCTGGAATGAAGAACAGCGCTGT 1095
QY 995 GGGATTGCTAGTTATCCCTCTTACCCAGAAATCT 1028
Db 1096 GGCATTGCCAACCTGGCCAGCTTCCCAAGATGT 1129

RESULT 3
PCT-US95-13820-1
; Sequence 1, Application PC/TUS9513820
; GENERAL INFORMATION:
; APPLICANT: Kneprl Pharmaceuticals, Inc.
; TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13820
; FILING DATE: 26-OCT-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US UNKNOWN
; FILING DATE: 02-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/330,121
; FILING DATE: 27-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: FP-60261-1-PC/DJR/RMS
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1482 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
```

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 142..1128

PCT-US95-13820-1

Query Match 27.5%; Score 310; DB 5; Length 1482;

Best Local Similarity 62.1%; Pred No 1e-85;

Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5.

```
QY 105 CTGGATCATCATCTGGAATCTCTGGAAGAAACCTCAGCAACAAT-CAAGCAAGAGAAAT 163
Db 205 CTGGACACCCACTGGGAGCTATGGAAGAAAGACCCAGAGAAACAATAA-AAAGATG 264
QY 164 GAGGAAGTAGCAGCGCTCTCATCTGGGAAAAAATCTAAAAATTTGTGATGCTTCACAAT 223
Db 265 GATGAATCTCTCGCGGTTTAAATTTGGGAAAAAACCCTCAAGATATTTCCATCCATAAC 324
QY 224 CTGGAACACTCAATGGGAATGCATCATATGATCTTAGGATTAAGATTAAGAAATTAAG 283
Db 325 CTGTAGGCTTCTCTTGGTCTCCATACATATGAACCTGGCTATGAACCCACCTGGGGACATG 384
QY 284 ACTGCTCAAGAAAGTGATATCTTTGATGGGTTCCCTCAGAGTTCCTTCTTCTTCTTCT 340
Db 385 ACCAGTGAAGAGGTGGTTTCAGAAAGATGACTGAACTCAAAAGTACCTCTCTCTCTCTCT 444
QY 341 AGAAATGTCACTCTTATAGCTCAAACTCTTAATCAGAAAATTCCTGATTTCTGTGGACTGG 397
Db 445 AGTAATGACACCTTTATATCCAGAAATGGGAAGGTAGAGCCCCACACTGTGTGACTAT 504
QY 398 AGAGAGAAGGGCTGTGTACTGAAGTGAATACCAAGGTTCTTGTGGTGTGTGTGGGCT 457
Db 505 CGAAAGAAAGCATATGTACTCTGTCAAAATACAGGGTCAAGTGTGTGTGTGTGTGTGTGT 564
QY 458 TTCAGCGCTGTGGGGCCCTGGAAAGCACAACCTCAAGCTAAAAACAGAAAGCTGTGTCT 517
Db 565 TTTAGTCTGTGGTGGCTGGAGGCGCACTCAAGNAGAAACTGCAAACTCTTAAAT 624
QY 518 CTGAGTGCACAGAACTTGTGTAGATTTGCTCAACTGMAAAATATGGGAATAAAGCTGCAAT 577
Db 625 CTGAGTCCCAAGCACTAGTGGATTTCTGTCTCTCA-----GAATGATGCTGTGGA 675
QY 578 GGGGTTTTCATGACAACTGCTTTCAGTATATTATTAATACAAAGCCATGATTCAGAA 637
Db 676 GGGGCTTACATGACCAATGCTTCCATATATGTCAGAGAAAGCCGGGTATTGACTCTGAA 735
QY 638 GCTTCTATCCCTACAAAGCCATGAATGGAAAGTGCAGATATCACTCAAAAAAGCGAGCT 697
Db 736 GATGCTTACCCATATGTGGGACAGAGAGAGTGTATGTACAAACCCCAACAGCAAGGCA 795
QY 698 GCCACATGTTCAAGATATCTGAACTTCCCTTTGGCAGTGAAGATGCTTAAAGAGAGCT 757
Db 796 GCTAAATGCAGAGGTACAGAGATATCCCGAGGGGAATGAGAAGCCCTCAAGAGGCA 855
QY 758 GTGCCAATAAAGGACCTGTGTGTGTGTATAGATTAAGATTAAGTATTTCTTCTGT 817
Db 856 GTGCCCGAGTGGGACCTGTCTGTGTGCGCAATGACAAAGCTTGAATCTCTCTCTCTCT 915
QY 818 TACAGAACTGTGTCTACTATGAACCATCTG---TACTCAGAAATGCAATCAGAGTA 874
Db 916 TACAGCAAGGTGTGATTAATGATGAAGTGTGCAATAGCAATAATCTGAACCATGCGGTT 975
QY 875 TTAGTGTGTGCTATGTTAACTTTAATGGGAAGACTACTGGCTTGTGAAAAACAGCTGG 934
Db 976 TTGCAGTGGCATATGGAATCCAGAGGGAACAAAGCAGCTGGATAATTAATAAACAGCTGG 1035
QY 935 GGCCTCAACTTGTGTGACCAAGGATATATCGGATGCAAGAACAGCTGGAAATCACTGT 994
Db 1036 GGAGAAACTGGGAAACAAAGGATATATCTCTATGCTGGAATGAAGAACAGCGCTGT 1095
QY 995 GGGATTGCTAGTTATCCCTCTTACCCAGAAATCT 1028
Db 1096 GGCATTGCCAACCTGGCCAGCTTCCCAAGATGT 1129
```


83 CTGGACACCCACTGGGAGCTATGGAAGAAGACCCACAGGAAGCAATATAACAACAAGGTG 142
164 GAGGAAGTAGCACAGGCTCTATCTGCGGAAAAAAATCTAAAATTGTGATGCTTCACAAT 223
143 GATGAATCTCTCGGCGTTAATTGGGAAAAAAACCTGAAGTATATTTCATCCATAAC 202
224 CTGGAACATCAATGGGAATGATATATGATCTAGGCATCAACCATCTGGGACATG 283
203 CTGAGGCTCTCTGGTCTCATACATGAATGAGTATGAACCATCTGGGACATG 262
284 CTGGTGAAGAGTATATCTTTGATGGTTCCTGAGAGTTCCC---AGCCAATGGCAG 340
263 ACCAGTGAAGAGTATGTTACTCTCTCAAAAATCAAGGTCAGTGGTTCCTGTTGGGCT 442
341 AGAATGTGCAC---TTATAGTGCAAACTCTAATCAGAAATTCCTGATTCCTGTCGACTGG 397
323 AGTAATGACACCTTTATATCCAGATGGGAAGTACAGCCCAAGACTCTGTCACAT 382
398 AGAGAGAAGGCTGTGTACTGAATGAATACAGGGTCTTGTGGTCTGTTGGGCT 457
383 CGAAGAAAGAGTATGTTACTCTCTCAAAAATCAAGGTCAGTGGTTCCTGTTGGGCT 442
458 TTGAGGCTGTGGGCGCTTGAAGCACAACCTGAAGTCAAAACAGGAAGCTGCTGTCT 517
443 TTGAGTCTGTGGTGCCTGAGAGTATGTTACTCAAAAATCAAGGTCAGTGGTTCCTGTTGGGCT 502
518 CTGAGTGCACAGCACTTTGATGCTCACTGAAAAATATGCGAATGAAGGCTGCAAT 577
503 CTGAGTCCCAAGCACTTAGTGGATGTTGTTCTCTGA-----GAATGATGCTGCA 553
578 GCGGCTTTCATGACAACTGCTTCCAGTATATTATTGATAAACAAGGCAATGATTCAGAA 637
554 GGGGCTTACATGACCAATGCTTCCCAATATGTCAGCAAGAACCGGGGTATGACTCTGAA 613
638 GTTCTCTATCCCTACAAAGCCATGAATGGAAGTGCAGATATGACTCAAAAAGCCAGCT 697
614 GATGCTATCCCATATGTGGGACAGGAGTGTATGTATCAAAACCAAGCAAGGCA 673
698 GCCACATGTTCAAGTATGCTGACCTTCCCTTTGGCAGTCAACATGCTTTAAAGAAGCT 757
674 GCTAAATGACAGGGTACAGAGATGCCCCGAGGGAATGAGAAAGCCCTGCAAGCGCA 733
758 GTGGCCAAATAAGGAGCTGTGCTGTGGTATAGATGCGGACCACTATTTCTTCTGCTG 817
734 GTGGCCGAGTGGGACCTGTCTCTGTGGCATTGATGCAAGGCTGACCTCTCTCCAGTTT 793
818 TACAGAAGTGTGTCTACTATGAACCATCTCTG---TACTCAGAAATGTGAATCATGGAGTA 874
794 TACAGCAAGTGTGTCTATTATGATGAAGCTGCAATAGCGATATCTGAACCATGCGGTT 853
875 TTAGTGTGTGCTATGTAAGCTTTAATGGGAGAGTACTGGCTTGTGAAAAACAGCTGG 934
854 TTGGCACTGGGATATGGAATCCAGAGGGAACAGCACTGGATTAATTAAGAAACAGCTGG 913
935 GGCCTCAACTTTGCTGACCAAGGATATATCGGATGCAAGAAACACTGGAATCACTGT 994
914 GGAGAAACTGGGAAACAAAGGATATATCTCATGCTCGAATTAAGAAACAGCTGT 973
995 GGGATGTGCTATTCCTCTTACCCAGAAATCT 1028
974 GGCATTGCCAACCTGGCCAGCTTCCCAAGATGT 1007

Query Match 27.5%; Score 310; DB 4; Length 1619;
Best Local Similarity 62.1%; Pred. No. 1.le-85;
Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;
QY 105 CTGATCATCTACTGGAATCTCTGGAAGAAAAACCTACAGCAACAAT-CAAGGAAGAGAAAT 163
Db 83 CTGGACACCCACTGGGAGCTATGGAAGAAGACCCACAGGAAGCAATATAACAACAAGGTG 142
QY 164 GAGGAAGTAGCACAGGCTCTATCTGCGGAAAAAAATCTAAAATTGTGATGCTTCACAAT 223
Db 83 CTGGACACCCACTGGGAGCTATGGAAGAAGACCCACAGGAAGCAATATAACAACAAGGTG 142
QY 143 GATGAATCTCTCGGCGTTAATTGGGAAAAAAACCTGAAGTATATTTCATCCATAAC 202
Db 224 CTGGAACATCAATGGGAATGATATATGATCTAGGCATCAACCATCTGGGACATG 283
QY 203 CTGAGGCTCTCTGGTCTCATACATGAATGAGTATGAACCATCTGGGACATG 262
Db 284 CTGGTGAAGAGTATATCTTTGATGGTTCCTGAGAGTTCCC---AGCCAATGGCAG 340
QY 263 ACCAGTGAAGAGTATGTTACTCTCTCAAAAATCAAGGTCAGTGGTTCCTGTTGGGCT 442
Db 341 AGAATGTGCAC---TTATAGTGCAAACTCTAATCAGAAATTCCTGATTCCTGTCGACTGG 397
QY 323 AGTAATGACACCTTTATATCCAGATGGGAAGTACAGCCCAAGACTCTGTCACAT 382
Db 398 AGAGAGAAGGCTGTGTACTGAATGAATACAGGGTCTTGTGGTCTGTTGGGCT 457
QY 383 CGAAGAAAGAGTATGTTACTCTCTCAAAAATCAAGGTCAGTGGTTCCTGTTGGGCT 442
Db 458 TTGAGGCTGTGGGCGCTTGAAGCACAACCTGAAGTCAAAACAGGAAGCTGCTGTCT 517
QY 443 TTGAGTCTGTGGTGCCTGAGAGTATGTTACTCAAAAATCAAGGTCAGTGGTTCCTGTTGGGCT 502
Db 518 CTGAGTGCACAGCACTTTGATGCTCACTGAAAAATATGCGAATGAAGGCTGCAAT 577
QY 503 CTGAGTCCCAAGCACTTAGTGGATGTTGTTCTCTGA-----GAATGATGCTGCA 553
Db 578 GCGGCTTTCATGACAACTGCTTCCAGTATATTATTGATAAACAAGGCAATGATTCAGAA 637
QY 554 GGGGCTTACATGACCAATGCTTCCCAATATGTCAGCAAGAACCGGGGTATGACTCTGAA 613
Db 638 GTTCTCTATCCCTACAAAGCCATGAATGGAAGTGCAGATATGACTCAAAAAGCCAGCT 697
QY 614 GATGCTATCCCATATGTGGGACAGGAGTGTATGTATCAAAACCAAGCAAGGCA 673
Db 698 GCCACATGTTCAAGTATGCTGACCTTCCCTTTGGCAGTCAACATGCTTTAAAGAAGCT 757
QY 674 GCTAAATGACAGGGTACAGAGATGCCCCGAGGGAATGAGAAAGCCCTGCAAGCGCA 733
Db 758 GTGGCCAAATAAGGAGCTGTGCTGTGGTATAGATGCGGACCACTATTTCTTCTGCTG 817
QY 734 GTGGCCGAGTGGGACCTGTCTCTGTGGCATTGATGCAAGGCTGACCTCTCTCCAGTTT 793
Db 818 TACAGAAGTGTGTCTACTATGAACCATCTCTG---TACTCAGAAATGTGAATCATGGAGTA 874
QY 794 TACAGCAAGTGTGTCTATTATGATGAAGCTGCAATAGCGATATCTGAACCATGCGGTT 853
Db 875 TTAGTGTGTGCTATGTAAGCTTTAATGGGAGAGTACTGGCTTGTGAAAAACAGCTGG 934
QY 854 TTGGCACTGGGATATGGAATCCAGAGGGAACAGCACTGGATTAATTAAGAAACAGCTGG 913
Db 935 GGCCTCAACTTTGCTGACCAAGGATATATCGGATGCAAGAAACACTGGAATCACTGT 994
QY 914 GGAGAAACTGGGAAACAAAGGATATATCTCATGCTCGAATTAAGAAACAGCTGT 973
Db 995 GGGATGTGCTATTCCTCTTACCCAGAAATCT 1028
QY 974 GGCATTGCCAACCTGGCCAGCTTCCCAAGATGT 1007
Db

RESULT 6
US-08-798-096-1
Sequence 1, Application US/08798096
Patent No. 6387682
GENERAL INFORMATION:
APPLICANT: Haslind, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PFI07D2
CURRENT APPLICATION NUMBER: US/08/798,096
CURRENT FILING DATE: 1997-02-12

Tue Jul 1 14:40:54 2003

PCT-US94-04781-1

Query Match 27.5% Score 310; DB 5; Length 1619;
Best Local Similarity 62.1%; Pred. No. 1.1e-85;
Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;
QY 105 CTGGATCATCTGGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAAT 163
DB 83 CTGGACACCTGAGGCTATGGAAGACACCCAGGAGCAATATAACAACAAGGTG 142
QY 164 GAGGAAGTAGACGGCTCTGCTGGAAGAAATCTAAATTTGTGATGCTTCAAT 223
DB 143 GATCAATCTCTGGCGCTTTAAATTTGGGAAAAAATCTGAAGTATATTTCCATCCATAAC 202
QY 224 CTGGAACACCTCAATGGGAATGCATTCATATGCTAGGCATGAACCATCTGGGAGACATG 283
DB 203 CTGAGGCTTCTCTGCTGCTCCATACATATGAATGCTATGAACCACTGGGGAGACATG 262
QY 284 ACTGCTGAAGAGTATATCTTTGATGGTTCCTGAGAGTTCCC---AGCCAATGGCAG 340
DB 263 ACCACTGAAGAGTCTGCTGAGAAGTACTGGACTCAAGTACCCCTGCTCATTCCTCCCGC 322
QY 341 AGAATGTCAC---TTATAGTCAAACTCTAATCAGAAATTTGCTGATTTCTGTGGACTGG 397
DB 323 AGTAATGACACCTTTATATCCCAAGATGGGAAGTAGAGCCCGAGACTCTGCTGACTAT 382
QY 398 AGAGAGAGGGCTGTGTTACTGAACTGAATACCAGGGTCTTGTGCTGCTTGTGGCT 457
DB 383 CCAAGAAAGGATATCTTACTCTCTGCAAAATCAGGCTCAGTGTGCTGCTGCTGGCT 442
QY 458 TTCAGGCTGTGAGGCTCTGAGACCAACTGAAGCTTAAACAGGAAAGCTGGTGTCT 517
DB 443 TTTAGCTGTGGGTGCTCTGGAGGCCAATCAAGAGAAACTGCGCAAACTCTTAAT 502
QY 518 CTGAGTGCACAGAACTTGGTATGATGCTCAACTGAAATATGGAATAAGAGCTGCAAT 577
DB 503 CTGAGTCCCAAGCTAGTGGATTGTGTCTGA-----GAATGATGGGTGTGGA 553
QY 578 GCGGCTTTCATCAGCACTGCTTTCCAGTATATTTGATGAACACGGCATTTGATTCAGAA 637
DB 554 GGGGCTACATGACCAATGCTTCCAAATATGTCAGAGAGAACCCGGGTATTCACTCTCAA 613
QY 638 GUTTCCTATGCTTACAAAGCCATGAATGAAAGTGCAGATATGACTCAAAAAGGAGCT 697
DB 614 GATGCTTACCATATCTGGACAGGAGAGTGTGATGTCAGAACCCCAAGCAAGGCA 673
QY 698 GCCACATGTTCAAGTATCTGAACCTTCCCTTTGGCAGTGAAGATGCTTAAAGAGAGCT 757
DB 674 GCTAATGACAGAGGTACAGAGATCCCGGAGGGAATGACAAAGCCCTGAAGAGGCA 733
QY 758 GTGGCCCAATAAAGAGCTGTGCTGTGGCTATAGTGGAGCCACTATTTCTTCTCCTG 817
DB 734 GTGGCCGAGTGGGAGCTGTCTGTGGCAATGATGCAAGGCTGACCTTCTTCAGTTT 793
QY 818 TACAGAGTGGTGTCTACTATGAACCATCTG---TACTCAGAATGTGAATCATGGAGTA 874
DB 794 TACAGAAAGGTGTGATATGATGAACCTGCAATAGCCATAACTGAACCATGGGTT 853
QY 875 TTAGTGTGTGGCTATGTAACCTTTAATGGGAAAGACTACTGCTTGTGGAACACAGCTGG 934
DB 854 TTGGCAGTGGATATGAATCCAGAGGGAACAAGCACTGGTAATAATTAAACACAGCTGG 913
QY 935 GGCCTCAACTTTGTTGACCAAGGATATTTCCGATGCAAGAAACAGTGAATCACTGT 994
DB 914 GGAGAAACTGGGAAACAAGGATATATCTCATGCTGCAATTAAGAACACAGCTGT 973
QY 945 GAGATTGCTATCTTATCTCTTTACCCAGAAATCT 1028
DB 974 GCATTTGCCAACCTGGCAGCTTCCCAAGATGT 1007

RESULT 9

US-08-964-308-3

; Sequence 3, Application US/08964308
; Patent No. 6066715
; GENERAL INFORMATION:
; APPLICANT: DESMARAIS, SYLVIE
; APPLICANT: FRIPSEN, RICHARD
; APPLICANT: ZAMBONTI, ROBERT
; TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE
; TITLE OF INVENTION: BINDING ASSAY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBERT J. NORTH - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,308
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: NORTH, ROBERT J
; REGISTRATION NUMBER: 27,366
; REFERENCE/DOCKET NUMBER: 19840 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-7262
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-964-308-3

Query Match 27.5% Score 310; DB 3; Length 1669;
Best Local Similarity 62.1%; Pred. No. 1.1e-85;
Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;
QY 105 CTGGATCATCTGGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAAT 163
DB 193 CTGGACACCACTGGGAGCTATGGAAGAAAGACCCAGGAGCAATATAACAACAAGGTG 252
QY 164 GAGGAAGTAGCAGCGGCTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCAAT 223
DB 253 GATCAATCTCTCGGCTTTAATTTGGGAAAAAACCCTCAAGTATATTTCCATCCATAAC 312
QY 224 CTGGAACATCAATGGGAATGCATTCATATGCTAGGCATGAACCATCTGGGAGACATG 283
DB 313 CTGAGGCTTCTCTGTGGTGTCCATACATATGAAGCTGCTATGACCACTCTGGGAGCATG 372
QY 284 ACTGCTCAAGAAAGTATATCTTTGATGGGTTCCCTCAGAGTTCCC---AGCCAATGGCAG 340
DB 373 ACCAGTGAAGAGTGGTTTCAGAAAGTACTGGACTCAAAAGTACCCCTGCTCATTCGCCG 432
QY 341 AGAATGTCAC---TTATAGTCAAACTCTAATCAGAAATTTGCTGATTCGTAATGCTG 397
DB 433 AGTAATGACACCTTTTATATCCCAAGATGGGAAGTGAAGCTGAGCCCACTCTGCTGACTAT 492
QY 398 AGAGAGAGGGCTGTGTTACTGAAATGAAATACAGGTTCTTGTGATGCTTGTGAGCT 457
DB 493 CGAAGAAAGGATATGTTACTCTCTCAAAATCAAGGCTCAGTGTGCTGCTGCTGGCT 552
QY 458 TTCAGGCTGTGGGGCCCTCGAAGACCACTGAAGCTAAACAGGAAAGCTGGTGTCT 517
DB 553 TTTAGCTCTGTGGTGTGCTGAGAGTAAATTTAAAGAAACTGCAACTCTTAAT 612

QY 875 TTAGTGGTGGCTATGCTAACTTAATGGAAGAACTACTGGCTTGTGAAAACACAGCTGG 934
DB 964 TTGCGAGTGGGATGGAATCCAGAGGGAACAGCACTGGATTAATTAAGAAACAGCTGG 1023
QY 935 GGCCTCAACTTTGCTGACCAAGGATATATTCGGATGGCAAGAAACAGTGGAAATCACTGT 994
DB 1024 GGAGAAACAGTGGGGAACAAAGGATATATCTCTCATGCTCGAAATAAGAACACAGCTGT 1083
QY 995 GGGATTGCTAGTTATCCCTCTTACCCAGAAATCT 1028
DB 1084 GGCAATGCCAACCTGGCCAGCTTCCCAAGATGT 1117

RESULT 11
US-09-069-138-3
; Sequence 3, Application US/09069138
; Patent No. 6348572
; GENERAL INFORMATION:
; APPLICANT: DESMARIS, SYLVIE
; APPLICANT: DUFRESNE, CLAUDE
; APPLICANT: FRIESEN, RICHARD
; APPLICANT: LEBLANC, YVES
; APPLICANT: ROY, PATRICK
; APPLICANT: YOUNG, ROBERT N.
; APPLICANT: ZAMBONI, ROBERT
; TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE
; TITLE OF INVENTION: BINDING ASSAY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PHILIPPE L. DURETTE - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Diskette
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/069,138
; FILING DATE: 29-APR-1998
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: DURETTE, PHILIPPE L.
; REGISTRATION NUMBER: 35,125
; REFERENCE/DOCKET NUMBER: 19840Y1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-4568
; TELEFAX: 732-594-4720
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-069-138-3

Query Match 27.5%; Score 310; DB 4; Length 1669;
Best Local Similarity 62.1%; Pred. No. 1.1e-85;
Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

QY 105 CTGGATCATCTGCAATCTCTGGAAGAAACCTTACAGCAAACT-CAAGGAAGAGAAAT 163
DB 193 CTGGACACCCACTGGAGCTATGGAGAGACCACAGGAAGCAATATAACAACAAGGTG 252
QY 164 GAGGAAGTAGCACGCGCTCTCATCTGGGAAAAAATCTAAATTTGTGCTTCACAAT 223
DB 253 GATGAATCTCTCGCGCTTTAATTTGGGAAAAAACCCTGGAAGTATATTTCCATCCATAAC 312

QY 224 CTGAACACACTCAATGGGAATGCAATTCATATGATCTAGGCATGAACATCTGGGACACATG 283
DB 313 CTTGAGGCTTCTCTGTTGCTCCATACATATGAATCTGGCTATGAACACCTGGGACACATG 372
QY 284 ACTGCTCAAGAGTGTATCTTTGATGGTTCCTCAGAGTTTCCC---AGCCAATGGCAG 340
DB 373 ACCAGTGAAGAGTGGTTTCAAGAGATGCTGGAGCTCAAGTACCCCTGTCTCATTTCCGC 432
QY 341 AGAAATGTCAC---TTATAGTCAAACTCTAAATCAGAAATTCCTGATTTCTGTGGACTGG 397
DB 433 AGTAATGACACCCCTTTATATCCAGATGGGAAGTAGAGCCGCCACACTCTGTGCACTAT 492
QY 398 ACAGAGAAGGCTGTGTTTACTGAAGTCAAAATACAGGTTCTTCTGCTGCTGCTTGGCT 457
DB 493 CGAAAGAAAGGATATGTTTCTCTGTCAAAATCAGGTCAGTGTGTTCTCTTGGCT 552
QY 458 TTCAGCGCTGTGGGGCCCTGGAAGCACAACTGAAGCTAAAACAGGAAAGCTGGTGTCT 517
DB 553 TTTAGCTCTGTGGTGGCTGGAGGGCAACTCAAGAGAAACTGGCAAACTCTTAAAT 612
QY 518 CTGAGTGCACAGAACTTGGTAGATTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAT 577
DB 613 CTGAGTCCCAAGACCTAGTGTGTTCTCA---GAATGATGCTGTGGA 663
QY 578 GCGGTTTTCATGACAACTGCTTTCAGTATATTAATTAACAACGCACTGATTCAGAA 637
DB 664 GGGGCTACATGACCAATGCTTCCATATGTCAGAGAAACCCGGGTATTTGACTCTGAA 723
QY 638 GCTTCTATCCCTACAAAGCCATGAATGGAAGTGCAGATATGACTCAAAAACCCAGCT 697
DB 724 GATGCTACCCATATGTCGGACAGAGAGACTTCTATGTACAACCCCAAGCAAGGCA 783
QY 698 GGCACATGTTCAAAGTATACTGAACTTCCCTTTGGCAGTGAAGTGCCTTAAAGAGCT 757
DB 784 GCTAAATGCAGAGGGTACAGAGATGCTCCCGAGGGAATGAGAAAGCCCTGAAGAGGCA 843
QY 758 GTGGCAATAAAGGACCTGTGCTGTGCTATAGATGCGAGCCACTATTCTTCTCTG 817
DB 844 GTGGCCGAGTGGGACCTCTCTGTGGCATTGATGCAAGCTTACAGCTTCTCTCTT 903
QY 818 TACAGAAGTGGTGTCTACTATGAACCATCTG---TACTCAGATGTCATCATGAGCTA 874
DB 904 TACAGCAAGGCTGTGTATTATGATGAAGCTCCCAATACGATAATCTCAACTATGCTGT 963
QY 875 ITAGTGTGGCTATGTTAAGCTTTAATGGGAAAGATTAAGCTTGTGAAAACAGCTGG 934
DB 964 TTGCGAGTGGGATATGGAATCCAGAGGAAACACAGCACTGGATAATTAAGAAACAGCTGG 1023
QY 935 GGCCTCAACTTTGGTGACCAAGGATATATTCGGATGGCAAGAAACAGTGGAAATCACTCT 994
DB 1024 GGAGAAACTGGGAAACAAAGGATATATCTCTGCTCGAAATAAGAAACAGCTTCT 1083
QY 995 GGGATTGCTAGTTATCCCTCTTACCCAGAAATCT 1028
DB 1084 GGCATTGCCAACCTGGCCAGCTTCCCAAGATGT 1117

RESULT 12
US-08-964-308-12
; Sequence 12, Application US/08964308
; Patent No. 6066715
; GENERAL INFORMATION:
; APPLICANT: DESMARIS, SYLVIE
; APPLICANT: FRIESEN, RICHARD
; APPLICANT: ZAMBONI, ROBERT
; TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE
; TITLE OF INVENTION: BINDING ASSAY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBERT J. NORTH - MERCK & CO., INC.
; STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
; CITY: RAHWAY


```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-964-313-12

Query Match 27.4% Score 308.4; DB 3; Length 990;
Best Local Similarity 62.0%; Pred. No. 2.7e-85;
Matches 579; Conservative 0; Mismatches 336; Indels 19; Gaps 5;

QY 105 CTGGATCATCACTGAATCTCTGGAGAAACCTCAGCAACAAT-CAAGGAAGAGAAT 163
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 CTGGACACCCATCGGAGCTATGGAAGAACCCACAGGAAGCAATATACACAAGCTG 123
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 164 GAGGAAGTAGCAGCGCTCTCATCTGGGAAAAAATCTAAAATTTGTGATGCTTCAAT 223
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 GATGAATCTCTCGCGTTAATTTGGGAAAAAACCTGAAGTATATTTCCATCCATAAC 183
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 224 CTGGAACACTCAATGGGAATGCATTCATATGATCTAGGCATGAACCATCTGGCAGACATG 283
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 CTGAGGCTCTCTTGGTGTCCATACATATGATGAAATGCTAATGAAATGAAATGAAATG 243
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 284 ACTGTTGAAGAAGTGTATCTTTGATGGTTCCCTGAGAGTTCC-AGCCAATGGCAG 340
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 ACCAGTGAAGAGTGTCTCAGAGATGATCTGAGCTCAAAATACCCCTGTCTCATTCGCCG 303
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 341 AGAATGTGAC---TTATAGTCAACTCTAATCAGAAATTCCTGATCTGTGACTGG 397
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 AGTAATGACACCTTTATATCCAGAAATGGGAAGTAGAGCCCGAGACTGTGCTGACTAT 363
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 398 AGAGAAAGGCTGTCTTACTGAAGTGAATACCAAGGTTCTTTGGTGGTGGCT 457
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 CGAAGAAAGGATATGTTACTCTCTCAAAATCAGGCTCAGTGTGGTCTCTTGGGCT 423
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 458 TTGAGGCTGTGGGCGCTTGAAGCACAACCTGAAGCTTAAACAGGAAGCTGTGCT 517
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 424 TTTAGCTCTGTGGTGGCTGGAGGCCCACTCAAGAAAGAACTGGCAAACTCTTAAT 483
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 518 CTGAGTCACAGAACTTGTGTAGATTCTCAACTGAAATATGGAATTAAGGCTGCAAT 577
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 484 CTGATCCCAAGACCTAGTGGATTGTGTCTGA-----GAATGATGGCTGTGA 534
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 578 GGGGTTTCATGACAACTGCTTCCAGTATATTTATGATTAACAAGGCTTATTCAGAA 637
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 535 GGGGCTTACATGACCAATGCTTCCAAATATGTCAGAAAGACCGGGGTATGACTCTGAA 594
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 638 GCTTCCTATCCCTACAAGCCTGAATGGAAGTGCAGATATGACTCAAAAAGCGAGCT 697
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 595 GATCCCTACCATATGTGGCAGAGAGAGAGAGAGTGTATGTACAAACCAAGGAGCA 654
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 698 GCCACATGTTCAAAGTATACTGAACCTTCCCTTTGGCAGTCAAGATGCCCTTAAAGAGCT 757
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 655 GCTAATGTCAGAGGTACAGAGATCCCGAGGGGAATGAGAAAGCCCTGAAGAGGCA 714
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 758 GTGGCAATAAAGACCTGTCTGTGGCTATAGATCCGAGCAGCTATTCTTTTCCTG 817
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 715 GTGGCCCGAGTGGAGCTGTCTGTGGCCATTTGATGCAAGCTGACCTCTTCCAGTTT 774
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 818 TACAGAGTGGTGTCTACTATGAACCAATCCTG---TACTCAGAAATGTGATCATGAGTA 874
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 775 TACAGAAAGTGTGTATATGATGAAGCTGCAATAGCCATATCTGAACCATGGGTT 834
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 875 TTATGGTGTGCTATGTTAAGCTTAATGGGAAGACTACTGGCTTTGTGAAAAACAGCTGG 934
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 835 TTGGCAGTGGGATATGGAATCCAGAGGGAACCAAGCACTGGATATTAANAACAGCTGG 894
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 935 GGCCTCAACTTTGTTGACCAAGGATATATTCGGATGGCAAGAACAGTGGAAATCACTGT 994
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 895 CGAGAAACTGGGAAACAAAGGATATATCTCTGCTGGAATGAAGAAACAAAGCGCTGT 954
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 995 GGAATGCTATGATCTCTTACCCAGAAATCT 1028
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 955 GGCATGGCAACCTGGCCAGGCTTCCCCAAGATGT 988
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

GENERAL INFORMATION:
APPLICANT: DESMARAIS, SYLVIE
APPLICANT: DUFRESNE, CLAUDE
APPLICANT: FRIESEN, RICHARD
APPLICANT: LEBLANC, YVES
APPLICANT: ROY, PATRICK
APPLICANT: ZAMBONI, ROBERT
APPLICANT: YOUNG, ROBERT N.
TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE
TITLE OF INVENTION: BINDING ASSAY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: PHILIPPE L. DURETTE - MERCK & CO., INC.
STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
CITY: RAHWAY
STATE: NJ
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/069,138
FILING DATE: 29-APR-1998
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: DURETTE, PHILIPPE L.
REGISTRATION NUMBER: 35,125
REFERENCE/DOCKET NUMBER: 19840Y1A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-4568
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 990 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-069-138-12

Query Match 27.4% Score 308.4; DB 4; Length 990;
Best Local Similarity 62.0%; Pred. No. 2.7e-85;
Matches 579; Conservative 0; Mismatches 336; Indels 19; Gaps 5;

QY 105 CTGGATCATCACTGGAATCTCTGGAAGAAACCTCAGCAACAAT-CAAGGAAGAGAAT 163
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 64 CTGGACACCCATCGGAGCTATGGAAGAACCCACAGGAAGCAATATACACAAGCTG 123
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 164 GAGGAAGTAGCAGCGCTCTCATCTGGGAAAAAATCTAAAATTTGTGATGCTTCAAT 223
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 GATGAATCTCTCGCGTTAATTTGGGAAAAAACCTGAAGTATATTTCCATCCATAAC 183
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 224 CTGGAACACTCAATGGGAATGCATTCATATGATCTAGGCATGAACCATCTGGCAGACATG 283
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 184 CTGAGGCTCTCTTGGTGTCCATACATATGATGAAATGCTAATGAAATGAAATGAAATG 243
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 284 ACTGTTGAAGAAGTGTATCTTTGATGGTTCCCTGAGAGTTCC-AGCCAATGGCAG 340
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 244 ACCAGTGAAGAGTGTCTCAGAGATGATCTGAGCTCAAAATACCCCTGTCTCATTCGCCG 303
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 341 AGAATGTGAC---TTATAGTCAACTCTAATCAGAAATTCCTGATCTGTGACTGG 397
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 304 AGTAATGACACCTTTATATCCAGAAATGGGAAGTAGAGCCCGAGACTGTGCTGACTAT 363
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 398 AGAGAAAGGCTGTCTTACTGAAGTGAATACCAAGGTTCTTTGGTGGTGGCT 457
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 CGAAGAAAGGATATGTTACTCTCTCAAAATCAGGCTCAGTGTGGTCTCTTGGGCT 423
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 458 TTGAGGCTGTGGGCGCTTGAAGCACAACCTGAAGCTTAAACAGGAAGCTGTGCT 517
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 424 TTTAGCTCTGTGGTGGCTGGAGGCCCACTCAAGAAAGAACTGGCAAACTCTTAAT 483
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 518 CTGAGTCACAGAACTTGTGTAGATTCTCAACTGAAATATGGAATTAAGGCTGCAAT 577
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 484 CTGATCCCAAGACCTAGTGGATTGTGTCTGA-----GAATGATGGCTGTGA 534
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 578 GGGGTTTCATGACAACTGCTTCCAGTATATTTATGATTAACAAGGCTTATTCAGAA 637
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 535 GGGGCTTACATGACCAATGCTTCCAAATATGTCAGAAAGACCGGGGTATGACTCTGAA 594
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 638 GCTTCCTATCCCTACAAGCCTGAATGGAAGTGCAGATATGACTCAAAAAGCGAGCT 697
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 595 GATCCCTACCATATGTGGCAGAGAGAGAGAGTGTATGTACAAACCAAGGAGCA 654
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 698 GCCACATGTTCAAAGTATACTGAACCTTCCCTTTGGCAGTCAAGATGCCCTTAAAGAGCT 757
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 655 GCTAATGTCAGAGGTACAGAGATCCCGAGGGGAATGAGAAAGCCCTGAAGAGGCA 714
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 758 GTGGCAATAAAGACCTGTCTGTGGCTATAGATCCGAGCAGCTATTCTTTTCCTG 817
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 715 GTGGCCCGAGTGGAGCTGTCTGTGGCCATTTGATGCAAGCTGACCTCTTCCAGTTT 774
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 818 TACAGAGTGGTGTCTACTATGAACCAATCCTG---TACTCAGAAATGTGATCATGAGTA 874
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 775 TACAGAAAGTGTGTATATGATGAAGCTGCAATAGCCATATCTGAACCATGGGTT 834
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 875 TTATGGTGTGCTATGTTAAGCTTAATGGGAAGACTACTGGCTTTGTGAAAAACAGCTGG 934
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 835 TTGGCAGTGGGATATGGAATCCAGAGGGAACCAAGCACTGGATATTAANAACAGCTGG 894
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 935 GGCCTCAACTTTGTTGACCAAGGATATATTCGGATGGCAAGAACAGTGGAAATCACTGT 994
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 895 CGAGAAACTGGGAAACAAAGGATATATCTCTGCTGGAATGAAGAAACAAAGCGCTGT 954
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 995 GGAATGCTATGATCTCTTACCCAGAAATCT 1028
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 955 GGCATGGCAACCTGGCCAGGCTTCCCCAAGATGT 988
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

QY 398 AGAGAGAAGGGCTGTGTTACTGAAGTGAATACACAGGGTTCTGTGTCGCTTTGGGCT 457
DB 364 CGAAGAAAGGATATGTTACTCTGCAAAATACAGGTCAGTGTGGTCTCTTGGGCT 423
QY 458 TTCAGGCTGTGGGGCCCTGGAGCACAACCTCAAGCTAAAAACAGGAAGCTGTGTCT 517
DB 424 TTAGCTCTGTGGTGGCTGGAGGACCACTCAAGMGAAGAACTGGCAACTCTTAAT 483
QY 518 CTGAGTCACAGAACTGTGTAGATCTCTCACTCACTGAATATATGGAATAAAGGCTGCAAT 577
DB 484 CTGAGTCCCAAGCAACTGTGTAGATCTCTCACTCACTGAATATATGGAATAAAGGCTGCA 534
QY 578 GAGGATTTTCATGACAACTGTCTTCCAGTATATTTATGATAACAAAGGCACTTGCATCAGAA 637
DB 535 GGGGCTACATGACCAATGCTTCCAAATATCTGCAGAACAGCGGGTATTGACTCTGAA 594
QY 638 GCTTCCTATCCCTACAAAGCCATGAATGAAAGTGCAGATATGACTCAAAAGCGAGCT 697
DB 595 GATGCTACCCATATGTGGACAGGAGAGAGTGTGTATGACAAAGCCCAAGGCAAGGCA 654
QY 698 GCCACATGTTCAAAGTATACTGAACCTTCCCTTTGGCAGTGAAGATGCCCTTAAAGAGCT 757
DB 655 GCTAAATGACAGGCTACAGAGATCCCGAGGGGAATGAGAAAGCCCTGAAGAGGCA 714
QY 758 GTGGCCAAATAAGGAGCTGTGTCTGTGTGTATAGATGAGGCACTATTTCTTCTGCTG 817
DB 715 GTGGCCGAGTGGGACCTGTCTGTGGCCATTTGATGCAAGCCTGACCTCTTCCAGTTT 774
QY 818 TACAGAAAGTGTGTCTACTATGACCATCTG---TACTCAGAACTGTAATCATGAGTA 874
DB 775 TACAGAAAGTGTGTATATGATGAAAGCTGCATAGGCAATATCTGAACCATCGGTTT 834
QY 875 TTAGTGTGTGCTATGCTAACTTTAATGGAAAGACTACTGCTTGTGAAAAACAGCTGG 934
DB 835 TTGGCAGTGGGATATGGAATCCAGAGGCAAGCAAGCACTGCTGATAATTAAGCAAGCTGG 894
QY 935 GGCCTCAATTTGTAACCAAGATATATTCGGATGGCAAGAAACAGTGGAAATCACTGT 994
DB 895 GGAGAAAGCTGGGAAACAAAGGATATATCTCATGCTCGAAATGAAGCAACAGCTGT 954
QY 995 GGATTGCTAGTATCCCTCTTACCAGAAATCT 1028
DB 955 GGTATTGCAACTGAGCTAGCTTCCCAAGATGT 988

RESULT 15

US-08-964-308-13
Sequence 13, Application US/08964308
Patent No. 6066715
GENERAL INFORMATION:
APPLICANT: DESMARIS, SYLVIE
APPLICANT: FRIESEN, RICHARD
APPLICANT: ZAMBONI, ROBERT
TITLE OF INVENTION: NEW LIGANDS FOR PHOSPHATASE
TITLE OF INVENTION: BINDING ASSAY
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: ROBERT J. NORTH - MERCK & CO., INC.
STREET: 126 EAST LINCOLN AVENUE - P.O. BOX 2000
CITY: RAHWAY
STATE: NJ
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,308
FILING DATE: 04-Nov-1996
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:

NAME: NORTH, ROBERT J
REGISTRATION NUMBER: 27,366
REFERENCE/DOCKET NUMBER: 19840 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 732-594-7262
TELEFAX: 732-594-4720
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 990 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-964-308-13

Query Match 27.2% Score 306.8; DH 3; Length 990;
Best Local Similarity 61.9% Pred. No. 8.3e-85;
Matches 578; Conservative 0; Mismatches 337; Indels 19; Gaps 5;
QY 105 CTGATCATCACTGGAATCTCTGGAAGAAACCTTACAGCAACAAAT-CAAGTAACACAAAT 163
DB 64 CTGGACACCCACTGGGAGCTATGGAAGAACACCCAGAGCAATATAA-AAACAAAGCTG 123
QY 164 GAGGAATAGTACGAGCGCTCTCATCTGCAAAAAAATCTAAATTTTGATG-TTAAAT 224
DB 124 CATGAATCTCTCGCGTTTAATTTGGAAAAAACCTTGAAGTATATTTCAACATAAC 183
QY 224 CTGGAACACTCAATGGGAATGCATTCATATGATCTAGGCATGAACCATCTGGGACATG 284
DB 184 CTGAGGCTCTCTGTGTGTCATACATATGAATGATGATGAAGCACTGGGAGCATG 243
QY 284 ACTGTTGAAGAAGTGCATATCTTTGATGGTTCCTCGAGAGTTCCTC---AGTCAATGGCAG 340
DB 244 ACCACTGAAGAGCTGGTTCAGAAATGACTGCACTCAAGATACCCCTGCTCATTCCTCCG 303
QY 341 AGAAATGTCA----TTATAGATCAAA-TTAATGAAATTTGGTCTATTCTGTGATG 397
DB 304 AGTAATGACACCTTTATATCCAGAA-TGGGAAGTACAGAGCCAGACTCTGTGCAAT 364
QY 398 AGAGAGAGGCTGTGTACTGAATCAATACAGGCTTCTGTGCTGTCTGTGCTGT 457
DB 364 GGAAGAAAGATATGTTATCTCTTAAATTAAGTCTAGTGTGTTGTTGTTGTTGTT 424
QY 458 TTCAGCCCTGTGGGGCCCTGGAGCACAACCTGAAGCTTAAATACGAAAGTGTGCTCT 517
DB 424 TTTAGCTCTGTGGTGGCTGGAGGCCCACTCAAGAAAGAAACTGCAAACTCTTAAAT 484
QY 518 CTGA-TG-A-A-A-A-TTGTGATGATG-TCAAT-TGAAATATATGGAATTAAGAGTGTAAAT 577
DB 484 CTGATCCCCAGAACCTAGTGGATTGTCTGCTGCA----GAATGATGATGTGTGGA 534
QY 578 GCGGTTTCATGACAACCTGCTTCCAGTATATTTATGAACAGCGCATGATTACAA 647
DB 535 GGGGCTACATGACCAATGCTTCCAAATATGTCAGAAAGAAAGCGGATATTGATCTGAA 594
QY 638 GCTTCTATCCCTACAAAGCCATGAATGGAAAGTGCAGATATGACTTAAAAAGCGAGCT 697
DB 595 CATGCTACCCATATGTGGGACAGAGAGAGTGTCTATGACAAACCAAGCAAGGCA 654
QY 698 GCCACATGTTCAAAGTATACTGAACTTCCCTTTGGCAGTGAAGATGCCCTTAAAGAGCT 757
DB 655 GCTAAATGACAGGCTACAGAGATCCCGAGGGGAATGAGAAAGCCCTGAAGAGGCA 714
QY 758 GTGGCCAAATAAGGAGCTGTGTGCTGCTATGATGCGAGGCACTATTTCTTCTGCTG 817
DB 715 GTGGCCGAGTGGGACCTGTCTGTGGCCATTTGATGCAAGCCTGACCTCTTCCAGTTT 774
QY 818 TACAGAAAGTGTGTCTACTATGACCATCTG---TACTCAGAACTGTAATCATGAGTA 874
DB 775 TACAGAAAGTGTGTATATGATGAAAGCTGCAATAGGCAATATCTGAACCATCGGTTT 834
QY 875 TTAGTGTGTGCTATGCTAACTTTAATGGAAAGACTACTGCTTGTGAAAAACAGTGG 934

Db 835 TTGGCAGTGGGATATGGATCCAGAGGGGAAACAGCAGTGGGATAATTAAAAACAGCTGG 894
QY 935 GGCTCAACTTTGGTGAACCAAGGATATATCGGATGGCAAGAAACAGTGGAAATCACTGT 994
Db 895 GGAGAAACTGGGGAAACAAAGGATATATCCTCATGGCTCGAAATAAGAAACAACGCCCTGT 954
QY 995 GGGATTCCCTAGTTATCCCTCTTACCCAGAAATCT 1028
Db 955 GGCATTGCTAACTGGCTAGCTTCCCAAGATCT 988

Search completed: July 1, 2003, 10:53:48
Job time : 81 secs


```
10b 241 AATGCAATTCATATGATCTAGGCATCAACCATCTGGAGACACAGCTGGTGAAGAAGTCAT 300
QY 301 ATCTTTGATGGGTCCCTCAGAGTCCAGAGCAATGGCAGAGAAATGTCACCTTATAGTGC 360
Db 301 ATCTTTGATGGGTCCCTCAGAGTCCAGAGCAATGGCAGAGAAATGTCACCTTATAGTGC 360
QY 361 AAACCTCAATCAGAAATTCCTGATTCCTGTGACCTGGAGAGAGGGGTGTGTACTGA 420
Db 361 AAACCTCAATCAGAAATTCCTGATTCCTGTGACCTGGAGAGAGGGGTGTGTACTGA 420
QY 421 AGTGAATACAGGGTTCCTGTGGTCTGTGTGGGCTTTCAGCGCTGTGGGGCCCTCGA 480
Db 421 AGTGAATACAGGGTTCCTGTGGTCTGTGTGGGCTTTCAGCGCTGTGGGGCCCTCGA 480
QY 481 AGCACAACCTGAAGCTAAAAACAGGAAGCTGTGTCTGTGAGTGCACAGAACTTGGTAGA 540
Db 481 AGCACAACCTGAAGCTAAAAACAGGAAGCTGTGTCTGTGAGTGCACAGAACTTGGTAGA 540
QY 541 TTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCGTTCATCACAACCTGCTTT 600
Db 541 TTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCGTTCATCACAACCTGCTTT 600
QY 601 CCAGTATATATGATAAGCAAGGCAATGATTCAGAAAGCTTCCTATCCCTACAAAGCCAT 660
Db 601 CCAGTATATATGATAAGCAAGGCAATGATTCAGAAAGCTTCCTATCCCTACAAAGCCAT 660
QY 661 GAATGAAAGTGCAGATATGACTCAAAAAGAGAGCTGCCACATGTTCAAAAGTATACTGA 720
Db 661 GAATGAAAGTGCAGATATGACTCAAAAAGAGAGCTGCCACATGTTCAAAAGTATACTGA 720
QY 721 ACTTCGCTTGGCAGTGAAGATGCCTTTAAAGAAAGCTGTGGCCAAATAAGGACCTGTGTC 780
Db 721 ACTTCGCTTGGCAGTGAAGATGCCTTTAAAGAAAGCTGTGGCCAAATAAGGACCTGTGTC 780
QY 781 TGTGGCTATAGATGGCAGCCTATCTTCTCTCTGTACAGAGTGGTGTCTACTATGA 840
Db 781 TGTGGCTATAGATGGCAGCCTATCTTCTCTCTGTACAGAGTGGTGTCTACTATGA 840
QY 841 ACATCCTCTACTCAGAAATGTAATCATGGAGTATTAGTGGTGGCTATGGTAACTTTAA 900
Db 841 ACATCCTCTACTCAGAAATGTAATCATGGAGTATTAGTGGTGGCTATGGTAACTTTAA 900
QY 901 TGGGAAGACTACTGGCTGTGCAAAACAGCTGGGGCTCAACTTTGGTGACCAAGGATA 960
Db 901 TGGGAAGACTACTGGCTGTGCAAAACAGCTGGGGCTCAACTTTGGTGACCAAGGATA 960
QY 961 TATTCGGATGGCAAGAACAGTGGAAATCACTGTGGATTTGCTAGTTATCCCTCTTACC 1020
Db 961 TATTCGGATGGCAAGAACAGTGGAAATCACTGTGGATTTGCTAGTTATCCCTCTTACC 1020
QY 1021 AGAAATCTAGCTCTCTATTTATATCANGTCCCAAAATTTGAACACCTTCTCTTAAT 1080
Db 1021 AGAAATCTAGCTCTCTATTTATATCANGTCCCAAAATTTGAACACCTTCTCTTAAT 1080
QY 1081 TTAATTTTACCTGCTATATAGTAAATAAATGATGTGATGACCAT 1126
Db 1081 TTAATTTTACCTGCTATATAGTAAATAAATGATGTGATGACCAT 1126
```

RESULT 2

```
US-10-099-275-1
: Sequence 1, Application US/10099275
: Publication No. US20020187499A1
: GENERAL INFORMATION:
: APPLICANT: Schneider, Patrick
: APPLICANT: Yamamoto, Karen K.
: APPLICANT: French, Cynthia K.
: APPLICANT: Reprogen, Inc.
: TITLE OF INVENTION: Use of Calhopsin S in the Diagnosis and Treatment of
: FILE OF INVENTION: Endometriosis
: CURRENT APPLICATION NUMBER: US/10/099,275
: CURRENT FILING DATE: 2002-04-13
```

```
: PRIOR APPLICATION NUMBER: US/09/701,685
: PRIOR FILING DATE: 2001-03-26
: PRIOR APPLICATION NUMBER: US 60/088,017
: PRIOR FILING DATE: 1998-06-04
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 1643
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (70)..(1065)
: OTHER INFORMATION: human cathopsin S
US-10-099-275-1
```

```
Query Match 75.1%; Score 845.4; DB 9; Length 1643;
Best Local Similarity 87.2%; Pred. No. 1e-240;
Matches 986; Conservative 0; Mismatches 136; Indels 9; Gaps 5;
```

```
QY 1 GCATTATTAGTATGGAGGACCTGCGAGTCTTCTTC---ATGAAATGGCTATTTGGGTGC 57
Db 32 GGAUCTTTATGTAAGAGCAAC-TG-TG-TT-TATTA-AAATGAAAGGCTGTGTGTGTC 91
QY 58 TTCCGTTTCTGCTCCTATGCGACAGTACATAAAGATGCCACTCTCGATCATCTACT 117
Db 92 TCTTGGTGTGCTCCTCTGCACTGCAACAGTTGCATTAAGATCTACTGATCATCTACT 151
QY 118 GGAATCTCTGCAAAANATATCAAAAT---AAGAGAGAGATAGAGAGATAGCAC 176
Db 152 GGCATCTCTGGAAGAAACCTATGCCAAACAAATACAAAGAAACAAATCAAGAACACTAC 211
QY 177 GGCCTCTCATCTGGAAGAAATATATAATTTTGTGATGATCA---AATCTGCAATCTCA 236
Db 212 GACCTCTCATCTGGAAGAAATCTAAGTTTGTGATGCTTCAACCTGGAGCATTCAA 271
QY 237 TGGGAATGCAATTCATATGATCTAGGCATGAACCATCTGGGAGACATGACATGGCAAG 296
Db 272 TGGGAATGCACTATAGCATCTGGCATGAACCACTTGGAGACATGACCAAGTGAAGAG 331
QY 297 TGATATCTTTGATGGGTTCCTGAGAGTTCGCCAGCAATGAGCAAAATGCAATTTATA 356
Db 332 TGATCTTTGATGAGTTCCCTGAGAGTTCGCCAGCAATGAGCAAAATGCAATTTATA 391
QY 357 GGTCAAACTCTAATCAGAAATTCGCTGATTCCTGGACTGGAGAGAGAGAGTGTGTTA 416
Db 392 AGTCAAACTCTAATCAGAAATTCGCTGATTCCTGGACTGGAGAGAGAGAGTGTGTTA 451
QY 417 CTGAAGTGAATACAGAGGTTCTTGTGCTGCTTCTTGGCTTTCAGCGTTCGGGGCCC 476
Db 452 CTGAAGTGAATACAGAGGTTCTTGTGCTGCTTCTTGGCTTTCAGCGTTCGGGGCCC 511
QY 477 TGGAGACAACTGGAAGCTAAACAGAGAAAGTGTGTTCTGTGATGTGAAGAAATTTGG 536
Db 512 TGGAGACAACTGGAAGCTAAACAGAGAGAGTGTGTTCTGTGATGTGAAGAAATTTGG 571
QY 537 TAGATTGCTCAACTGGAAGAAATATGAGAAATAAAGCTGCAATGAGAGGTTTTCATCAG 596
Db 572 TGGATTGCTCAACTGGAAGAAATATGAGAAATAAAGCTGCAATGAGAGGTTTTCATCAG 641
QY 597 CTTTCCAGTATATTATTGATAACAGCGCATTCATTGATCAGAGCTTCTCTATCTTACAA 656
Db 632 CTTTCCAGTATATTATTGATAACAGCGCATTCATTGATCAGAGCTTCTCTATCTTACAA 691
QY 657 CCATGAATGGAAGTGCAGATATGACTCAAAAAGAGAGAGTGGCACATGTTCAAAATATA 716
Db 692 CCATGAATGGAAGTGCAGATATGACTCAAAAAGAGAGAGTGGCACATGTTCAAAATATA 751
QY 717 CTGAAGTCTCTTGTGAGTGTGAATATTTAAANANANATGTTGAGAGAGAGAGAGTGT 776
Db 752 CTGAAGTCTCTTGTGAGTGTGAATATTTAAANANANATGTTGAGAGAGAGAGAGTGT 811
QY 777 TGTCTGTGCTATAGATGCCAGGACCATTTCTTTCTCTCTCTCTCTCTCTCTCTCTCT 836
```

```

Db      812  TGTCTGTTGGTGTACATGCGGCTATCTTCTTCTTCCCTACAGAGTGGTGTCTACT 871
Qy      837  ATGAACCATCTCTACTCAGAAATGTGAATCATGAGATTAATAGTGGTATGGTATGTAACC 896
Db      872  ATGAACCATCTCTACTCAGAAATGTGAATCATGAGATTAATAGTGGTATGGTATGTAACC 931
Qy      897  TTAATGGGAAGAGTACTGCGCTCTGAAAGAACAGCTGGGCGCTCAACTTTTGGTGAACCAAG 956
Db      932  TTAATGGGAAGAGTACTGCGCTCTGAAAGAACAGCTGGGCGCTCAACTTTTGGTGAACCAAG 991
Qy      957  CATATATTCGGATCGCAAGAACAGTGGAAATCACTGGGATTTAGTATTCCTCTTT 1016
Db      992  CATATATTCGGATCGCAAGAACAGTGGAAATCACTGGGATTTAGTATTCCTCTTT 1051
Qy      1017  ACCGAAATCTA---GACCTCTTCAATTTTATAAAGTCCAAATAAT-TGAACACATTT 1072
Db      1052  ACCGAAATCTAAGAGATCTCTCTTTTATAAATAATCAATGAATAATGAAGCACTTT 1111
Qy      1073  CTCTTAATTTAATTTTACCTGCTATA-ATAGTAAATAATAATTTGTCATGA 1122
Db      1112  CTCTTAATTTAATTTTACCTGCTATAAAGAAATAATTTGTCATGA 1162

```

```

RESULT 3
US-09-990-064-1
; Sequence 1, Application US/09990064
; Publication No. US20020164765A1
; GENERAL INFORMATION:
; APPLICANT: Parvinder K. Punia
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30242
; CURRENT APPLICATION NUMBER: US/09/990, 064
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: GB0028462.0
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Mammalia
US-09-990-064-1

```

```

Query Match      69.6%; Score 784; DB 9; Length 996;
Best Local Similarity 87.3%; Pred. No. 1.4e-222;
Matches 870; Conservative 0; Mismatches 125; Indels 1; Gaps 1;

Qy      36  ATGAAATGGCTAGTGGGCTGCTTCCGTTGCTCTCTATGCGAGTGGCAAGTACATAAA 95
Db      1  ATGAAATGGCTAGTGGGCTGCTTCCGTTGCTCTCTATGCGAGTGGCAAGTGGCAGCA 60
Qy      96  GATCCCACTCTGGATATCACTGGAATCTCTGGAAGAAACCTTACAGCAAAACANT-CAAG 154
Db      61  CACCTTACCCTGGATACCACTGGCATCTCTGGAAGAAACCTTACAGCAAAACANTCAAG 120
Qy      155  GAAGAGATAGGAAGTAGCAGGCGCTCATCTGGGAAATAATCTAAATTTGGATG 214
Db      121  GAAAGATAGGAAGTAGCAGGCGCTCATCTGGGAAATAATCTAAATTTGGATG 180
Qy      215  CTTCAATCTGGAACACTCAATGGGAATCATTCATCATGCGATGCAACCACTG 274
Db      181  CTTCAACCTGGAGCACTTCCATGGGAATGCACTGCTATGACGTGGCATGAACCACTG 240
Qy      275  GCAACATGATGTTGGTGAAGAGTATATCTTTGATGGGTTCCCTGAGAGTTCCAGGCAA 334
Db      241  CGACATGACCAAGCAAGAGAGTGGTGTCTTTGATGATGATGTTCCCTGAGAAATCCCAAC 300
Qy      335  TGGCAGAGAAATGTCATTTATAGTCAAACTCTAATCAGAAATTTGCTGATTTCTGTGGAC 394
Db      301  TGGCAGAGAAATGTCATTTATAGTCAAACTCTAATCAGAAATTTGCTGATTTCTGTGGAC 360
Qy      395  TCGAGAGAGAGAGGCTGTGTTACTGAAGTGAATACCAAGGCTTCTTGTGCTGTGTTGG 454

```

```

Db      361  TCGAGAGAGAGGGGCTGTGCTACTGAAGTCAAAATATACGGTCTTCTTGGTCTGCTCG 420
Qy      455  GCTTTACGCGCTGTGGGGCCCTGGAAAGCAAACTGAAGTAAACACAGAAAGCTGGTG 514
Db      421  GCTTTACGCGCTGTGGGGCCCTGGAAAGCAAACTGAAGTAAACACAGAAAGCTGGTG 480
Qy      515  TCTCTGAGTGCACAGAACTTGGTATGTTCAACTCAAAATATATGGGAATAAAGCTGC 574
Db      481  TCTCTGAGTGCACAGAACTTGGTATGTTCAACTCAAAATATATGGGAATAAAGCTGC 540
Qy      575  AATGCGGTTTTCATGACAACTGCTTCCAGTATATATTGATTAACAAAGGCTATGATCA 634
Db      541  AATGCGGTTTTCATGACAGAGGCTTCCAAATACATCATGATCAACAAAGGCTATGATCA 600
Qy      635  GAAGCTTCTATCCCTACAAAGCCATGAATGGAAGTGCAGATATGATCAAAAGAGCGA 694
Db      601  GAAGCTTCTATCCCTACAAAGCCATGATCAAAAGTGCAGATATGATCAAAAGAGCGA 660
Qy      695  GCTGCGCATGTTCAAAAGTATACACTTCCCTTTTGGAGTGAAGTGGCTTAAAGAA 754
Db      661  GCTGCGCATGTTCAAAAGTATACACTTCCCTTTTGGAGTGAAGTGGCTTAAAGAA 720
Qy      755  GCTGCGCATGTTCAAAAGTATACACTTCCCTTTTGGAGTGAAGTGGCTTAAAGAA 814
Db      721  GCTGCGCATGTTCAAAAGTATACACTTCCCTTTTGGAGTGAAGTGGCTTAAAGAA 780
Qy      815  CTGTACAGAAAGTGGTGTCTACTATGAACCATCTCTGCTACTCAGAAATGTAATCAT 874
Db      781  CTGTACAGAAAGTGGTGTCTACTATGAACCATCTCTGCTACTCAGAAATGTAATCAT 840
Qy      875  TTAGTGGTGGCTATGTTAACTTAAATGGGAAGACTACTGCTTGTCAAAAGAGCTGG 934
Db      841  CTAGCAGTGGCTATGTTAACTTAAATGGGAAGACTACTGCTTGTCAAAAGAGCTGG 900
Qy      935  GGCTTCAACTTTTGGTGACCAAGATATATTCGGATGCGCAAGAAACAGTGGAAATCAT 994
Db      901  GGCATCACTTCGGTGAACAGATATATTCGGATGCGCAAGAAACAGTGGAAATCAT 960
Qy      995  GGGATGCTAGTATCCCTCTTACCCAGAAATCTAG 1030
Db      961  GGGATGCTAGTATCCCTCTTACCCAGAAATCTAG 996

```

```

RESULT 4
US-10-010-577-1
; Sequence 1, Application US/10010577
; Publication No. US20030104971A1
; GENERAL INFORMATION:
; APPLICANT: thurmond, robin l
; APPLICANT: baker, sherry
; APPLICANT: karlsson, lars
; TITLE OF INVENTION: polynucleotide and polypeptide sequences of monkey cathepsin s
; FILE REFERENCE: ORT1457
; CURRENT APPLICATION NUMBER: US/10/010,577
; CURRENT FILING DATE: 2001-11-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 993
; TYPE: DNA
; ORGANISM: monkey
US-10-010-577-1

```

```

Query Match      68.2%; Score 768; DB 9; Length 993;
Best Local Similarity 87.0%; Pred. No. 7.9e-218;
Matches 867; Conservative 0; Mismatches 125; Indels 4; Gaps 2;

Qy      36  ATGAAATGGCTAGTGGGCTGCTTCCGTTGCTCTCTATGCGATGGGCAAGTACATAAA 95
Db      1  ATGAAGCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 60
Qy      96  GATCCCACTCTGGATCATCACTGGAATCTCTGGAAGAAACCTTACAGCAAAACAT-CAAG 154

```

```
Db 61 GATCCACCCCTGGATCACCACCTGGAATCTCTGGAAGAAAACCTACGCGCAACAATATAA 120
QY 155 GAAGACAATCAGGAAGTACGACGGGTCTCATCTCGGAAAAAATCTAAATTTGTGATG 214
Db 121 GAAAGATGAAGAAGACGATGACGCTCATCTCGGAGAGAATCTAAAGTTTGTGATG 180
QY 215 CTTCAACAATCTGGAACACTCAATGGATGCAATTCATATGATCTAGGCATGAACCATCTG 274
Db 181 CTTCAACAATCTGGAACACTCAATGGATGCAATTCATATGATCTAGGCATGAACCATCTG 240
QY 275 GGAGACATGATCTGGTGAAGAGTGAATATCTTTGATGGGTTCCCTGAGAGTTCGCCAGCAA 334
Db 241 GGAGACATGATCTGGTGAAGAGTGAATATCTTTGATGGGTTCCCTGAGAGTTCGCCAGCAA 300
QY 335 TGGCAGAGAAATGTCACTTATAGTCAAACTCTAATCAAAATTCGCTGATCTGTGCAC 394
Db 301 TGGCAGAGAAATGTCACTTATAGTCAAACTCTAATCAAAATTCGCTGATCTGTGCAC 360
QY 395 TGGAGAGAGAGGGCTGTCTTACTGAAGTGAATATACCAGGGTCTTTGTGGTCTTTGG 454
Db 361 TGGAGAGAGAGGGCTGTCTTACTGAAGTGAATATACCAGGGTCTTTGTGGTCTTTGG 420
QY 455 GCTTTCAGCGCTGTGGGGCCCTGGAAGCAGCAACTGAAGCTTAATAACAGGAAGCTGGT 514
Db 421 GCTTTCAGCGCTGTGGGGCCCTGGAAGCAGCAACTGAAGCTTAATAACAGGAAGCTGGT 480
QY 515 TCTCTCAGTGGCAGCAACTTTGATGCTCAACTGAAATATGGAATATGGAATATGGAATATG 574
Db 481 TCTCTCAGTGGCAGCAACTTTGATGCTCAACTGAAATATGGAATATGGAATATGGAATATG 537
QY 575 AATGGGGTTTATATCAAAATGCTTTTCCAGTATATATGATGAACAGGGCATGATCA 634
Db 538 AATGGGGTTTATATCAAAATGCTTTTCCAGTATATATGATGAACAGGGCATGATCA 597
QY 635 GAAGCTTCCTATCCCTACAAAGCCATGAATGAAAGTGCAGATATGACTCAAAAGCGA 694
Db 598 GAAGCTTCCTATCCCTACAAAGCCATGAATGAAAGTGCAGATATGACTCAAAAGCGA 657
QY 695 GCTGCGACATGTTCAAAAGTATCACTCAACTTCCCTTTGCGAGTGAAGATGCCCTTAAAGAA 754
Db 658 GCTGCGACATGTTCAAAAGTATCACTCAACTTCCCTTTGCGAGTGAAGATGCCCTTAAAGAA 717
QY 755 GCTGTGGCAATTAAGGACCTGTCTGTGGCTATAGATGCGAGCCACTATCTTCTTC 814
Db 718 GCTGTGGCAATTAAGGACCTGTCTGTGGCTATAGATGCGAGCCACTATCTTCTTC 777
QY 815 CTGTACAGAGTGGTGTCTTACTATGAACCATCTCTACTCAAGATGTGAATCATGGAGTA 874
Db 778 CTGTACAGAGTGGTGTCTTACTATGAACCATCTCTACTCAAGATGTGAATCATGGAGTA 837
QY 875 TTAGTGGTGGTATGGTAACTTAACTTAACTGGAAGACTACTGGCTTTGTAAGAAACAGCTGG 934
Db 838 CTGTGATGGTATGGTAACTTAACTTAACTGGAAGACTACTGGCTTTGTAAGAAACAGCTGG 897
QY 935 GGCCTCAACTTTGGTGACCAAGGATATATTCGGATGGCAAGAAACAGTGGAAATCACTGT 994
Db 898 GGCAGCAACTTTGGTGACCAAGGATATATTCGGATGGCAAGAAATTAAGAGAAACCACTGT 957
QY 995 GGGATTGCTAGTATCCCTCTTACCCAGAAATCTAG 1030
Db 958 GGGATTGCTAGTATCCCTCTTACCCAGAAATCTAG 993
```

RESULT 5

```
US-09-918-995-36579
; Sequence 36579, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
```

```
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36579
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-36579
```

```
Query Match 29 68; Score 333.8; DB 9; Length 417;
Best Local Similarity 87.5%; Pred. No. 3.5e-88;
Matches 365; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
```

```
QY 393 ACTGGAGACAGAGCGGTGTCTTACTGAAGTGAATACCAGGGTCTTTGCTGCTGCTTT 452
Db 1 ACTGGAGAGAGAAAGGGTGTCTTACTGAAGTGAATATCAAGGTCTTTGCTGCTGCTTT 60
QY 453 GGGCTTTCAAGCGCTGTGGGGCCCTGCAAGCACAACCTGAAGCTAAAAACAGCAAGCTGG 512
Db 61 GGGCTTTCAAGCGCTGTGGGGCCCTGCAAGCACAACCTGAAGCTAAAAACAGCAAGCTGG 120
QY 513 TGTCTCTGAGTGCACAGAACTTGTGATGATTTCTTAATGAAAAATATGGAATATAAAGCT 572
Db 121 TGTCTCTGAGTGCACAGAACTTGTGATGATTTCTTAATGAAAAATATGGAATATAAAGCT 180
QY 573 GCAATGGGGTTTTCATCACAACCTGCTTTCCAGTATATATGATAACAGCGATATTCATT 642
Db 181 GCAATGGTGGCTTTCATCACAACCTGCTTTCCAGTATATATGATAACAGCGATATTCATT 240
QY 633 CAGAAGCTTCTCTATCCCTACAAAGCCATGAATGCAAGTGCAGATATGATTTAAAAAGC 692
Db 241 CAGAAGCTTCTCTATCCCTACAAAGCCATGAATGCAAGTGCAGATATGATTTAAAAAGC 300
QY 693 GAGTGGCACATGTTCAAGTATATGATGATTTGCTTTGGCAGTGAAGATGCTTTAAAG 752
Db 301 GTGCTGGCACATGTTCAAGTATATGATGATTTGCTTTGGCAGTGAAGATGCTTTAAAG 360
QY 753 AAGCTGTGGCAATTAAGGACCTGTCTGTGGCTATAGATGCGAGCCACTATCTTCT 804
Db 361 AAGCTGTGGCTTAAAGAAAGCTGTCTGTGGCTATAGATGCGAGCCACTATCTTCT 417
```

RESULT 6

```
US-09-918-995-37273
; Sequence 37273, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 37273
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-37273
```

```
Query Match 29 48; Score 331.4; DB 9; Length 418;
Best Local Similarity 87.7%; Pred. No. 3.5e-88;
Matches 362; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
QY 520 GAGTGGACAGAACTTGTGATGATTTGCTCAACTGAAAAATATGGAATATAAAGCTTGAATGG 574
Db 6 GCGTGGCAGAACTTGTGATGATTTGCTCAACTGAAAAATATGGAATATAAAGCTTGAATGG 65
```

341 AGAATGTCAC---TTATAGTCAAACTCTATCAGAAATTCCTGATTCGTGGACTGG 397
323 AGTATGACACCCCTTTATATATCCAGAAATGGGAAGGTACAGCCCCACACTCTGTGCATAT 382
398 AGAGACAAGGGCTGTGTACTGAAGTGAATACCAAGGTTCTTGTGGTCTGTGGGCT 457
383 CGAAGAAGAGATATCTTACTCTGTCAAAAATCAGGCTAGTGTGGTCTGTGGGCT 442
458 TTCAGGGCTGTGGGGCCCTGGAACACAACTGAAGCTAAAACAGGAAGAGCTGTGTCT 517
443 TTTAGCTCTGTGGGTGCCCTGGAGGGCCAACTCAAGAGAACTGGCAAACTTTAAAT 502
518 CTGAGTGCACAGAACTTGGTAGATTGCTCAATGTAATAATATGGAATTAAGAGCTGCAAT 577
503 CTGAGTCCCAAGAACCTAGTGGATTGTGTCTGA-----GAATGATGGCTGTGGA 553
578 GGGGGTTTCATGACAACTGCTTTCCACTATATATTATGATAACAAGCGGATTCATTCAGAA 637
554 GGGGGCTTACATGACAACTGCTTTCCAAATATGTGCAGAGAACCGGGGTATTGATTCGAA 613
638 GCTTCCCTATCCCTACAAAGCCATGAATGGAAGCTGCAGATATCACTCAAAAAAGCGAGCT 697
614 GATCCCTACCCATATGTGGGACAGAGAGAGATTGTATGTACAACCCAAACAGGCAAGCA 673
698 GCCACATGTTCAAAGTATCTGAATCTTCCCTTTGGCACTGAAGATGCTTAAAGAGAGCT 757
674 GCTAAATCCAGAGGTACAGAGAGATCCCGGAGGGAATGAGAAAGCCCTGAAGAGGCA 733
758 GTGCCAATAAAGAGCTGTGTCTGTGCTATAGATCCGAGCCACTATTCTTTCTTCCTG 817
734 GTGGCCGAGTGGGACCTGTCTGTGGCCATTGATGCAAGCTGACCTCTCCCTCCAGTTT 793
818 TACAGAGTGGTGTCTACTATGAACCAATCCTG---TACTCAGAAATGTGAATCATGAGTA 874
794 TACACAAAGGTGTCTATTATGATGAAAGCTGCAATAGCCATATCTGAAACCATGCGGT 853
875 TTAGTGGTGGCTATGTGTAACCTTAATGGGAAGAGCTACTGGCTTGTGAAAAACAGCTGG 934
854 TTGGCAGTGGATATGGAATCCAGNAGGAACAGCACTGGATAATTTAAAAACAGCTGG 913
935 GGCCTCAACTTTGGTGACCAAGGATATATTCGGATGCAAGAAACAGTGGAAATCACTGT 994
914 GGAGAAAACCTGGGAAACAAAGGATATATCTCTGCTCGAAATTAAGAACACAGCTGT 973
995 GCGATTGCTAGTTATCTCTTCTTACCCAGAAATCT 1028
974 GCGATTGCCAACCTGGCCAGCTTCCCAAGATGT 1007

RESULT 8
US-10-114-464-1
; Sequence 1, Application US/10114464
; Patent No. US20020142448A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D5
; CURRENT APPLICATION NUMBER: US/10/114,464
; CURRENT FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: 08/553,125
; PRIOR FILING DATE: 1995-11-07
; PRIOR APPLICATION NUMBER: 08/208,007
; PRIOR FILING DATE: 1994-03-08
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (21)..(1009)
; FEATURE:

580 CAGTTTATGACAACTGCTTTCCAGTATATATATGATAAACAAGGCAATGATTGAGAGC 639
66 TGGCTTCATGACAAGGGCTTTCCAGTACATCAATGATAACAAGGCAATGATTGAGAGC 125
640 TTAATATGACAACTGCTTTCCAGTATATATATGATAAACAAGGCAATGATTGAGAGC 699
126 TTCTATCCCTACAAAGCCATGGATCAGAAATGTCAATATGACTCAAAATATCGTGTGC 185
700 CACATGTTCAAAGTATGAACTTCCCTTTGGGAGTGAAGATGCTTTAAAGAGAGCTGT 759
186 CACATGTTCAAAGTATGAACTTCCCTTTGGGAGTGAAGATGCTTTAAAGAGAGCTGT 245
760 GGCCAATAAAGAGGCTGTGTCTGTGGCTATAGATGCGAGCCACTATTCTTTCTCCTGTA 819
246 GGCCAATAAAGAGGCTGTGTCTGTGGCTATAGATGCGAGTGTCAATCATGAGATATTAGT 879
820 CAGAAAGTGGTGTCTACTATGACCACTCTGTGCTGATGAGATGGAATGGAATGGAATGGA 305
306 CAGAAAGTGGTGTCTACTATGACCACTCTGTGCTGATGAGATGGAATGGAATGGAATGGA 879
880 GCTTGGCTATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 932
366 GCTTGGCTATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGA 418

Query Match 27.5%; Score 310; DB 10; Length 1619;
Best Local Similarity 62.1%; Pred. No. 1.9e-81;
Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;
105 CTGATGATCATCTGGAATCTCTGGAAGAAACCTACAGCAACAAAT-CAAGCAAGAGAT 163
83 CTGACACCCACTGGGAGCTATGGAAGAGAGCCACAGGAAGCAATATAACAAGAGTG 142
164 GAGAGATGACAGGCTGTCTATCTGGAAGAAATCTAAAAATTTGTGATGCTTCAAT 223
143 GATGAAATCTCTGGGCTTTAAATTTGGGAAACAAACCTGAGTATATTTCCATCCATAAC 202
224 CTGAGACATCAATGGAATGCAATTCATATGATCTAGGATGCAATCTGGGAGCATG 283
203 CTGAGGCTCTCTTGGTGTCCATACATATGATGATGATGATGATGATGATGATGATGATG 262
284 ACTGCTGAGCAAGTATGATCTTTGATGGTTCCTCAGAGTTCCC---AGCCATGGCAG 340
263 ACCAGTGAAGAGTGGTTCAGAGATGATGAGTGGACTCAAGATACCCCTGCTCATTCCTCCG 322

US-09-953-956-1
; Sequence 1, Application US/09953956
; Patent No. US20020072107A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PF107D5D1
; CURRENT APPLICATION NUMBER: US/09/953,956
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: 09/219,441
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (21)..(1009)
; NAME/KEY: mat_peptide
; LOCATION: (365)..(1009)
; NAME/KEY: sig_peptide
; LOCATION: (21)..(1009)
US-09-953-956-1

NAME/KEY: mal_peptide
LOCATION: (365)..(1009)
FEATURE:
NAME/KEY: sig_peptide
LOCATION: (21)..(1009)
US-10-114-464-1

Query Match 27.5% Score 310; DB 12; Length 1619;

Best Local Similarity 62.1% Pred. No. 1.9e-81;

Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

QY 105 CTGGATCATCA-TGG-AAT-TGTGGAAGAAAACTTA-AGTAAAAT-CAAGGAAGAGAAAT 163
DB 83 CTGGACACCAAT-TGG-AAG-TATGGAAGAGAAAGAAACAGGAAGCAATATAACAACAAGGTG 142
QY 164 GAGGAAGTAGACCGGCTCACTCGGGAAGAAATCTAAATTTGTGATGCTTCACAAAT 223
DB 143 GATGAATCTCTCGCGGTTTAATTTGGGAAAAAACCTGAAGTATATTTCCATCCATAAC 202
QY 224 CTGGACACCTTAAT-TGG-AAT-TATGATATATATGAGCATGAACCATCTGGGACACATG 283
DB 203 CTTGAGGCTTCTCTGGTGCTCCATACATGAACCTGGCTATGAACCACTGGGACATG 262
QY 284 ACTGTTGAAGAGTGNATCTTTGATGGTTCCCTGAGAGTTCC-AGCCAATGGCAG 340
DB 263 ACCAGTGAAGAGGTGTTTCAGAGATGACTGAGCTCAAGTACCCCTGTCTCAITCCCGC 322
QY 341 AGAAATGTCAC---TTATAGTCAAACTCTAATCAGAAATTCCTGATTTCTGTGACATGG 397
DB 323 AGTATGACACCTTTATATCCAGNATGGGAGGTAGAGCCAGAGCTCTGTGACATAT 382
QY 398 AGAGAGAGGGCTGTGTACTGAAGTGAATACCAAGGTTCTTTGTGGTCTTTGGGCT 457
DB 383 CGAAGAAAGGATATGTTACTCTCTCAAAAATCAGGCTCACTGTGTTCTCTTTGGGCT 442
QY 458 TTCAGCGCTGTGGGGCCCTGGAAGCAACCTGAAGCTAAACAGAGGAAGCTGTGTCT 517
DB 443 TTTAGCTCTGTGGTCCCTGGGGCCCACTCAAGAGAAACCTGGCAAACTCTTAAT 502
QY 518 CTGAGTGCACAACTTGGTAGATTGCTCAACTGAAAAATATGGAATAAAGGCTTGCAT 577
DB 503 CTGAGTCCCAAGCACTAGTGATTGTGTGCTGA-----GAATGATGGCTTGA 553
QY 578 GCGGTTTTCATGCAACCTCTTCCAGTATATTATGATACAGGCAATGATTCAGAA 637
DB 554 GGGGCTACATACCAATGCCCTTCCAATATGTCAGAGAAACCGGGGTATTCACTCTGAA 613
QY 638 GCTTCTATCCCTACAAAGCCATGAATGCAAGTGTAGATGACTCAAAAAGGAGCT 697
DB 614 GATGCTACCCATATGTGGACAGAGAGAGATTGTATGATACACCCCAACAGGCA 673
QY 698 GGCACATGTTCAAACTACTGAACCTTCCCTTTGGCAGTGAAGATGCCCTTAAAGAGCT 757
DB 674 GCTAAATGCAAGGTTACAGAGAGATCCCGAGGGAATGAGAAAGCCCTCAAGAGGCA 733
QY 758 GTGKCAATAAAGAGCTGTGTCTGTGCTATAGATGAGCCACTATTTCTTCTCTG 817
DB 734 GTGGCCGAGTGGGACCTGTCTCTGTGGCCATTTGATCAAGCCCTGACCTCTCTCAAGTT 793
QY 818 TACAGAGTGGTGTCTACTATGAACCATCCTG---TACTCAGATGTGAATCATGGAGTA 874
DB 794 TACAGCAAGGTGTGTATTATGATGAAGCTGCAATAGTGATAATCTGAACCATCGGTT 853
QY 875 TTAGTGGTGTGCTATGTTAACTTAACTTAATGGGAAGACTACTGGCTTGTGAAAAACAGCTG 934
DB 854 TTGGGAGTGGGATATGGAATCAGAGGGAACAGCACTGATTAATTAACCAAGCTGG 913
QY 935 GGCCTCAACTTTGATGACCAAGATATATTCGATGGCAAGAAACAGTGGAAATCACTGT 994
DB 914 GAGAAACTGGGGAACCAAGGATATATCTCTCATCTGCTCGAATAAGAACACCCCTGT 973
QY 995 GATATGTTAGTTATGCTCTTACCCAGAAATCT 1028

DB 974 GGCATTGCCAACCTGGCCAACTTTCUCCUANAAGT 1007

RESULT 9

US-10-044-090-462

: Sequence 462, Application US/10044090

: Patent No. US20020137081A1

: GENERAL INFORMATION:

: APPLICANT: Olga Bandman

: TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION

: FILE REFERENCE: PA-0028 US

: CURRENT APPLICATION NUMBER: US/10/044.090

: CURRENT FILING DATE: 2002-01-09

: NUMBER OF SEQ ID NOS: 850

: SOFTWARE: PERL Program

: SEQ ID NO 462

: LENGTH: 1671

: TYPE: DNA

: ORGANISM: Homo sapiens

: NAME/KEY: misc-feature

: OTHER INFORMATION: Inocyte ID No. US20020137081A1 4114137H1

US-10-044-090-462

Query Match 27.5% Score 310; DB 12; Length 1671;

Best Local Similarity 62.1% Pred. No. 2e-81;

Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

QY 105 CTGCATCATCAGTCGGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAAT 163
DB 169 CTGGACACCACTCTGGAGCTATGGAAGAGAGAGCCACAGGAAGCAATATAACAACAAGGTG 228
QY 164 GAGCAAGTAGCACCGGCTCTCATCTGGCAAAAAAATCTAAATTTGTGATGCTTCACAAAT 223
DB 229 GATCAATCTCTCGCGCTTAAATTTGGAAAAAACCTGAAGTATATTTCCATCCATAAC 298
QY 224 CTGGAACACTCAATGGGAATGATTCATATGATCTAGGATGAGCAATCTGGAAGAGATG 283
DB 289 CTTGAGGCTTCTCTTGGTGCTCCATACATATCAACTGGCTATGAACCACTGGGAGCATG 348
QY 284 ACTGCTCAAGAGTGTATCTTTGATGGTTCCCTGAGAGTTCGCT-AGCAATGCGAG 440
DB 349 ACCTATTAAGAGAGTGTTCAGAGATGATGAGCTCAAGTACGCTGTCTATTCTGCGC 408
QY 341 AGAATGTCAC---TTATAGTCAAACTCTAATCAGAAATTTGCTCATTTCTGGAATGG 497
DB 409 AGTAATCACACCCCTTATATCCAGAAATGGGAAGTAGAGCCCACTCTCTGCAATAT 468
QY 398 AGAGAGAAAGAGTGTGTATGTAAGTGAATATGAGAGTCTTGTGTTGTTCTGAGT 457
DB 469 CGAAGAAAGGATATGTTACTCTCTCAAAAATCAGGGTCACTGTGTTCTGTTGGCT 528
QY 458 TTCAGCCCTGTGGGGCCCTGGAGCACAACTCAAGCTTAAACAGCAAGCTGCTCTCT 517
DB 529 TTTAGTCTCTGAGTGTCTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 588
QY 518 CTGAGTGCACAGAACTTGGTAGATTGCTCAACTCAAAAATATGGGAATTAAGGTTGCAAT 577
DB 589 CTGAGTCCCAAGCAACCTAGTGGATTGTGCTCTGA-----GAATGATGCTCTGGA 634
QY 578 GAGGTTTTCATGCAACCTGCTTCCAGTATATTATGATACAGGCAATGATTCAGAA 637
DB 640 GGGGCTACATACCAATGCCCTTCCAATATGTCAGAGAAACCGGGGTATTGATCTGAA 694
QY 638 GCTTCTATCCCTACAAAGCCATCAATGGAAGTGGAGATATGACTCAAAAAGCTGAGT 697
DB 700 GATGCTTACCCATATGTTGGACAGGAGAGAGTGTATGATACACCCCAACAGGCA 754
QY 698 GGCACATGTTCAAGATATCTGAACTTCCCTTTGGGAGTGAAGATGCTTAAAGAGAT 757
DB 760 GCTAAATGCAAGGTTACAGAGATGCCCTGAGGGAAGTGAAGAGAGAGAGAGAG 814
QY 758 GTGGCAATAAAGAGCTTGTGTGTGATATATGAGGCAATATTTCTTCTCTGCTG 817

OTHER INFORMATION: n = A,T,C or G
US-10-198-846-10186

```
Query Match      27.5%; Score 310; DB 9; Length 2017;
Best Local Similarity 62.1%; Pred. No. 2.2e-81;
Matches 580; Conservative 0; Mismatches 335; Indels 19; Gaps 5;

QY 105 CTGGATCATCTGCAATCTCTGGAGAAACCTACAGCAAACT-CAAGCAAGAAAT 163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 192 CTGGACACCCACTGGAGGTATGGAGAGACCCACAGGAGCAATATACCAACAGGTG 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 164 GAGGAAGTACGACGGCTCTACTCTGGGAAAAAATCTAAATTTCTGATGCTTCACAA 223
    || ||| ||| ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
DB 252 GATGAAATCTCTCGCGCTTAATTTGGGAAAAAACCTCAACTATATTCCATCCATAAC 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 224 CTGGACACCTCAATGGAATGATTCATATGATCTAGCATGAACCATCTGGGACATG 283
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 312 CTGGAGCTTCTCTGCTGCTCATATGATGAACCTGGGTATGAACCACTGGGGACATG 371
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 284 ACTGGTGAAGAGTGTATCTTTTGATGGCTTCCCTCAGACTTCCC---AGCCAATGGCAG 340
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 372 ACCAGTGAAGAGTGTTCAGAAAGATGA-TGAA-TTAAAGTA-TCTGTCTTATTCCCGC 431
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 341 AGAAATGTCACT---TTATAGTCAAACTCTAATCAGAAATTCCTGATCTGTGACATGG 397
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 432 ACTAATGACACCTCTTATATCCAGAAATGGGAGGTAGAGCCGCCAGACTCTCTGCACTAT 491
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 398 AGAGAGAGGGCTGTCTTACTGAAGTGAATACACAGGGTCTTGTGGTCTTGTGGGCT 457
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 492 CGAAAGAAAGGATATGTTACTCTGTCAAAAATCAGGGTCACTGTGGTCTTGTGGGCT 551
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 458 TTACGGCTGTGGGGCCCTGGAAGCAACACTGAAGCTAAACAGAGGAAGCTGTGTCT 517
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 552 TTTAGCTCTGTGGGTGCCCTGGAGGCCAACTCAAGAGAAACTGGCAAACTCTTAAT 611
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 518 CTGAGTGCACAGAACTGTGATCTTCACTCAACTGAATAATATGGCAATAAAGCTGCAAT 577
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 612 CTGAGTCCCAAGCACTAGTGATGTGTGTCTGA-----GAATGATGGCTGTGA 662
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 578 GGGGTTTTCATGACAACTCTTCCAGTATATTTATGATAACAGGCAATGATTCAGAA 637
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 663 GGGGCTACATGACCAATCCCTTCCAATATGTGCAGAGAACCGGGGTATTGCACTTGAA 722
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 638 GCTTCTCTATCCCTACAAACCATGAATGAAAGCTGCAATATGACTCAAAAGCGAGCT 697
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 723 GATGCTTACCTCATATGTGTGA-TAG-AAGAG-A-TTGTATGTA-AA-TCCAA-AGGCAAGGCA 782
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 698 GCCACATGTTCAAAGTATATCTCAACTTCCCTTTGGCAGTGAAGATGCCCTTAAACAAAGCT 757
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 783 GCTAAATGCAGAGGTACAGAGATCCCGAGGGGAATGAGAAAGCCCTGAAGAGGCA 842
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 758 GTGGCAATAAAGAG-TGTGT-TGTGTATAGATGGAGG-TATATTCTTTCTTCCTG 817
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 843 GTGGCCGAGTGGGAGCTGTCTGTGGCCATTGATGCAAGCCCTGACCTCTCTCCAGTT 902
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 818 TACAGAGTGTCTCTACTATCAACCATCTC---TACTCAGAAATGTGAATCATGAGTA 874
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 903 TACAGCAAGGTGTGTATATATGAAAGTTCGAATAGCGAATATCTGAACCATCGGTT 962
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 875 TTAGTGGTTGGTATGTAACCTTAATGGGAAGACTACTGTGCTTGTGAAAAACAGCTGG 934
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 963 TTGGCAGTGGGATATGGAATCCAGAGGGGAAACAGCACTGGATATTTAAACAGCTGG 1022
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 935 GGGCTCACTTTGTGCAAGGATATATTCGGATGGCAAGCAAGCAAGTGGAAATCACTGT 994
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1023 GGAGAAACTGGGGAACAAAGGATATATCTTCATGCTGCTGCAATATGAAGCAACGCCGT 1082
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 995 GGGATTGTAGTTATCCCTCTTACCCAGAAATCT 1028
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1083 GGCATTGCAACTGGCTAGT-TTCCCAAGATGT 1116
    || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 12

```
US-09-960-352-13717
; Sequence 13717, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 13717
; LENGTH: 329
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 59-HOVMS1-011-Q1-E1-G12
US-09-960-352-13717

Query Match      24.1%; Score 271.4; DB 10; Length 329;
Best Local Similarity 89.1%; Pred No 2.1e-70;
Matches 293; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 665 GCAAGTCCACATATGACTCAAAAAGCGAGCTGCCACATCTTCAAAAGTATATGAACTT 724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 GGAAGTCCATATGATGTAAAAATGAG-TGACAGGTGTTCAAGTATATGAACTT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 725 CCCTTTGGCAGTGAAGATGCCCTTAAAGAGCGCTGGCCAAATAAGGACCTGTGTCTGTG 784
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 CCCTTGGCAGCAAGAGCGCTTAAAGAGAGCGCTGGCCAAATAAGGACCTGTGTCTGTG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 785 GCTATAGATGAGGAGTCACTATCTTCTCTCTATAGAGTGTGTCTACTATGAACTT 844
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 GGTATAGAGCGCAAGCTTCTTCTCTCTACAAAAGCTGGTGTCTACTATGACCTCC 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 845 TCCTGTACTCAGATGTGAATCATGAGTATTTAGTGGTGTGCTATGCTTAAGCTTAATGG 904
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 TCTCTA-T-A-T-AAGATGTGACATGCTAT-TCTGTGCTTGGTATGCTAAG-TTGAGGG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 905 AAACACTACTGCTTGTCAAAAAGCTGGGCGCTTCAACTTGGTGACCAATATATT 964
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 AAGCACTACTGCTTGTCAAAAAGCTGGGCGCTTCAACTTGGTGACCAATATATA 1000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 965 CGGATGGCAAGCAAACTGGAATCACTG 993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 CGGATGGCAAGCAAACTGGAATCACTG 329

RESULT 13
US-09-960-352-12526
; Sequence 12526, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12526
; LENGTH: 369
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 53-LIB34-084-Q1-E1-F2
US-09-960-352-12526

Query Match      22.3%; Score 250.8; DB 10; Length 369;
```


Search completed: July 1, 2003, 12:03:59
Job time : 232 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 09:22:27 ; Search time 2955 Seconds
(without alignments)
9580.522 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1126
Sequence: 1 gcatltagtgatgagaca.....aataatgtgtcatgaccat 1126

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243625 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_NA_Maln.*
1: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
8: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
9: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
10: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
11: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
12: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
13: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
14: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
15: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
16: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
17: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
18: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
19: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
20: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
21: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
22: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
23: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
24: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
25: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
26: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
27: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
28: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
29: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
30: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
31: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
32: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
33: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
34: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
35: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
36: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
37: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
38: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
39: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
40: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
41: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
42: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*
43: /cgn2_6/ptodata/1/pna/US086 COMB.seq.*

44: /cgn2_6/ptodata/1/pna/US6000 COMB.seq.*
45: /cgn2_6/ptodata/1/pna/US6001 COMB.seq.*
46: /cgn2_6/ptodata/1/pna/US6002 COMB.seq.*
47: /cgn2_6/ptodata/1/pna/US6003 COMB.seq.*
48: /cgn2_6/ptodata/1/pna/US6004 COMB.seq.*
49: /cgn2_6/ptodata/1/pna/US6005 COMB.seq.*
50: /cgn2_6/ptodata/1/pna/US6006 COMB.seq.*
51: /cgn2_6/ptodata/1/pna/US6007 COMB.seq.*
52: /cgn2_6/ptodata/1/pna/US6008 COMB.seq.*
53: /cgn2_6/ptodata/1/pna/US6009 COMB.seq.*
54: /cgn2_6/ptodata/1/pna/US6010 COMB.seq.*
55: /cgn2_6/ptodata/1/pna/US6011 COMB.seq.*
56: /cgn2_6/ptodata/1/pna/US6012 COMB.seq.*
57: /cgn2_6/ptodata/1/pna/US6013 COMB.seq.*
58: /cgn2_6/ptodata/1/pna/US6014 COMB.seq.*
59: /cgn2_6/ptodata/1/pna/US6015 COMB.seq.*
60: /cgn2_6/ptodata/1/pna/US6016 COMB.seq.*
61: /cgn2_6/ptodata/1/pna/US6017 COMB.seq.*
62: /cgn2_6/ptodata/1/pna/US6018 COMB.seq.*
63: /cgn2_6/ptodata/1/pna/US6019 COMB.seq.*
64: /cgn2_6/ptodata/1/pna/US6020 COMB.seq.*
65: /cgn2_6/ptodata/1/pna/US6021 COMB.seq.*
66: /cgn2_6/ptodata/1/pna/US6022 COMB.seq.*
67: /cgn2_6/ptodata/1/pna/US6023 COMB.seq.*
68: /cgn2_6/ptodata/1/pna/US6024 COMB.seq.*
69: /cgn2_6/ptodata/1/pna/US6025 COMB.seq.*
70: /cgn2_6/ptodata/1/pna/US6026 COMB.seq.*
71: /cgn2_6/ptodata/1/pna/US6027 COMB.seq.*
72: /cgn2_6/ptodata/1/pna/US6028 COMB.seq.*
73: /cgn2_6/ptodata/1/pna/US6029 COMB.seq.*
74: /cgn2_6/ptodata/1/pna/US6030 COMB.seq.*
75: /cgn2_6/ptodata/1/pna/US6031 COMB.seq.*
76: /cgn2_6/ptodata/1/pna/US6032 COMB.seq.*
77: /cgn2_6/ptodata/1/pna/US6033 COMB.seq.*
78: /cgn2_6/ptodata/1/pna/US6034 COMB.seq.*
79: /cgn2_6/ptodata/1/pna/US6035 COMB.seq.*
80: /cgn2_6/ptodata/1/pna/US6036 COMB.seq.*
81: /cgn2_6/ptodata/1/pna/US6037 COMB.seq.*
82: /cgn2_6/ptodata/1/pna/US6038 COMB.seq.*
83: /cgn2_6/ptodata/1/pna/US6039 COMB.seq.*
84: /cgn2_6/ptodata/1/pna/US6040 COMB.seq.*
85: /cgn2_6/ptodata/1/pna/US6041 COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1126	100.0	1126	38	US-10-010-580-1
2	856.4	76.1	2205	23	US-09-606-680-4112
3	856.4	76.1	2205	25	US-09-652-123-0177
4	856.4	76.1	2205	25	US-09-652-126-0929
5	856.4	76.1	2205	28	US-09-710-281-4558
6	856.4	76.1	2205	28	US-09-710-281-4558
7	856.4	76.1	2205	28	US-09-710-281-4558
8	856.4	76.1	2205	29	US-09-721-580-5798
9	856.4	76.1	2205	29	US-09-726-172-2249
10	856.4	76.1	2205	29	US-09-726-172-2249
11	856.4	76.1	2205	29	US-09-726-803-2191
12	856.4	76.1	2205	29	US-09-726-803-2191
13	856.4	76.1	2205	29	US-09-726-811-5142
14	856.4	76.1	2205	30	US-09-760-475-1251
15	856.4	76.1	2205	30	US-09-760-475-1251
16	856.4	76.1	2205	17	US-09-359-922-8469
17	856.4	76.1	2205	17	US-09-359-922-8469
18	856.4	76.1	2205	34	US-09-919-002-8469
19	856.4	76.1	2205	34	US-09-919-002-8469
20	856.4	76.1	2205	76	US-60-324-185-4019
21	856.4	76.1	2205	65	US-60-213-360-3680
22	856.4	76.1	2205	39	US-10-090-275-1

[illegible]

us-10-010-580-1.rnrm

Tue Jul 1 14:40:55 2003

	Db	704	TGGAAGCACACGCTGAAGCTGAAAAACGAAAGCTGGTGCTCTCTCATGTCCCAAGCCTGG	761
	Qy	537	TAGATTGCTCAACTCAAAAATATGGGAATAAAGGCTGCAATTGGGGTTTCATGCACAAC TG	596
	Db	762	TGGATTGCTCAACTGAAAAATATGGAACNAAGGCTGCAATGGTGGCTTCATGACAAAGG	821
	Qy	597	CTTTCCAGTATATTATGATTAACACGGCATTGATTCAAGAAGCTTCTGATCCCTACAAAG	656
	Db	822	CTTTCCAGTACATATTGATTAACAAGGCATCGACTCAGAGCCTTCCCTATCCCTACAAG	881
	Qy	657	COATGAANTGGAAGTCAGATATGACTCMAAAAAGCGAGCTGCCACATGTTTCAAAGTATA	716
	Db	882	CCAATGGATCAGAAATGTCNAATGACTCANAAATATCTGCTSCCAACATGTTCAAAAGTACA	941
	Qy	717	CTGAACCTTCCCTTTGGCAGTGAAGATGGCTTAAAAAGACGCTGTGCCCAATAAGGACCTG	776
	Db	942	CTGAACCTCCTTATGGCAGAGAAGATGTCTGAAAGAGCTGTGCGCAATAAAGGCGCCAG	1001
	Qy	777	TGCTCTGCTATAGATGGAGCCACTATTCTTCTCTGTACAGAAGTGCTGTCTACT	836
	Db	1002	TGFTCTGTGGTGTAGATGGCGTCACTCTTCTCTCTCTACAGAAGTGCTGTCTACT	1061
	Qy	837	ATGAACCATCCTGTACTCAGAAATGGAATCATGAGATATTAGTGGTGGCTATGSTATCC	896
	Db	1062	ATGAACCATCCTGTACTCAGAAATGGAATCATGCTGTACTTGTGGTGGCTATGSGTAGTC	1121
	Qy	897	TTAATGGGAAGACTACTGGCTGTCAAAAACAGCTGGGCCCTCAACTTTTGGTGACCAAG	956
	Db	1122	TTAATGGGAAGAAATACTGGCTTGTGAAAANACASCTGGGGCCACACTTTTGGTGAAGAAG	1181
	Qy	957	GATATATTCCGGATGGCAAGAAACAGTGGAAATCACTGTGGGATTTGCTAGTTATCCCTCT	1016
	Db	1182	GATATATTCCGGATGGCAAGAAATAAAGGAAATCATTTGTGGGATTTGCTAGCTTTCCCTCT	1241
	Qy	1017	ACCAGAAATCTA--GACCTCTTCATTTTATACAAAGTCCAAAAAATTCGAAACACTTTC	1073
	Db	1242	ACCAGAAATCTAGAGGATCTCTCCTTTTTATTACAAATCAAGNAATATGAAGCACTTTC	1301
	Qy	1074	TCCTTAATTTAATTTTACTCTCTAPA-ATAGTAAAAATAAATGTGTCATGA	1122
	Db	1302	TCCTTAATTTTCTGCTGTATCCAGGAAGAAATAAATTTGTGTCATGA	1351

RESULT 6

```

US-09-716-972-4397
; Sequence 4397, Application US/09716972
; GENERAL INFORMATION:
; APPLICANT: Hunter, John J.
; APPLICANT: Stylian, Andrew W.
; APPLICANT: Vasicek, Thomas
; APPLICANT: Leg, John
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600-2030-001
; CURRENT APPLICATION NUMBER: US/09/716,972
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/166,948
; PRIOR FILING DATE: 1999-11-22
; NUMBER OF SEQ ID NOS: 5251
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4397
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2205)
; OTHER INFORMATION: n = A,T,C or G
US-09-716-972-4397

```


	Matches	986;	Conservative	0;	Mismatches	136;	Indels	8;	Gaps	4;
Qy	1	GCATTATTA	TACTATGGGAGCACCTCGACGGTCTCTTTTC---	-	ATCAAAATGGCTAGTTGGGTGCG	57				
Ib	222	GGACTCTTA	CTGTGGGACCACTGCTGGTTCTATCACAATGAACGGCTGTTTGTTGTCGC	281						
Qy	58	TTCCGTTGT	GCTCTCATGAGTGGCACAAGTAGACATAAAGATCCCACCTCTGGATCATCATC	117						
Ib	282	TCATTGGTGT	GCTCTCTGCASTGGCAGATTGCAATAAAGATTCCTACCCCTGCATCACCAC	341						
Qy	118	GGAACTCTC	TGGAAGAAAACCTACAGCAAAACAAT--CAAGGAAGACAAATCAGCAAGCTAGCAC	176						
Ib	342	GGCATCTCT	GGAGAAAAACCTATGGCAACAACATACAAAGGAAGAATGAAGAAGCAGTAC	401						
Qy	177	GGCGTCTC	ATCTGGGAAAAAAATCTAAAAATTTGGTATGCTTTCACAAATCTGGAACACTCAA	236						
Ib	402	GAGCTCTCA	TCTGGGAAAAGAATCTAAAGTTTGTGATGCTTCACAAGCTGGAGCATTCAA	461						
Qy	237	TGGCAATGC	AATCTATATGCTTAGCGATGAACATCTGGAGATATCACTGGTGAATAAG	296						
Ib	462	TGGGAATGC	CACTACCGATCTGGGCATGAACCACTGGGAGACATGACCAGTGAAGAAG	521						
Qy	297	TGATATCTT	GTATGGGTTCCTGAGAGTTCCCAAGCCAATGGCAGAGAAATGTCTCATTTATA	356						
Ib	522	TGATGTCTT	TGATGAGTTCCTGAGAGTTGCCAGCCAGTGGCAGAGAAATATCACATATA	581						
Qy	357	GGTCAAACT	CTAATCAGAAATTCGCTGATTTCTGTGGACTGGAGACGAAGGGCTGCTTTA	416						
Ib	582	AGTCAAACT	CCTAATCGGATATTCGCTGATTTCTGTGGACTGGACACAAAGGGTGTGTTA	641						
Qy	417	CTGAAGTGA	AAATACCAAGGGTCTTTGTGGTGCCTTTGTGGGCTTTTCAGCGCTGTGGGGCCCC	476						
Ib	642	CTGAAGTGA	AAATATCAAGGTTCTGTGGTGCCTGGCTGGCTGGCTGGGGCCCC	701						
Qy	477	TGGAAGCAC	AACTGAACCTAAAACAGGAAGCTGCTCTCTCAGTGCACAGAACCTTGG	536						
Ib	702	TGAGGACAC	AGCTGAAGCTGAACAGGAAGCTGTGCTCTCAGTGCACAGAACCTTGG	761						
Qy	537	TAGATTGCT	CAACTGAAAAATATGGGAATAAAAGGCTGCAATGGCGGTTTTTCATGACAACTG	596						
Ib	762	TGGATTGCT	CAACTGAAAAATATGGAACAAAAGGCTGCAATGGTGGCTTCATGACAACTG	821						
Qy	597	CTTTCCAGT	ATATTATATANANGCAATTTGATTCAGAACTTCATATCCCTACAAAG	656						
Ib	822	CTTTCCAGT	ATATATGATAAAGGGCATCGACTCAGAGCTTCCTATCCCTTACAAAG	881						
Qy	657	CCATGAATG	GAAAGTGCAGATATCACTCAAAAAAGCGAGCTGGCACATGTTCAAAGTATA	716						
Ib	882	CCATGATCA	GAAATGTCAATATCACTCAAAATATCTGTGTGCACATGTTCAAGTACA	941						
Qy	717	CTCAACTTC	CCCTTTGGCAGTCAAGATGCCTTTAAAGAGCTGTGGCCAAATAAAGGACCTG	776						
Ib	942	CTCAACTTC	CCCTTTATGGCAGAGACATGCTCTGAAAGAGCTGTGGCCAAATAAAGGCCCCAG	1001						
Qy	777	TGCTTTGCT	ATATATGACAGCACTATTCTTTCTCTGTACAGAAAGTGGTGTCTACT	836						
Ib	1002	TGCTCTGT	TGTTAGATGCGCGTCACTCTTTCTTCTCTCTACAGAAAGTGGTGTCTACT	1061						
Qy	837	ATCAACCAT	CTCTACTCAGAATGTCAATCATGGAGTATTAGTGGTTGGCTATGGTAAC	896						
Ib	1062	ATGAACCAT	CTCTACTCAGAATGTGAATCATGGTGTACTTTGGTTGGCTATGGTGATC	1121						
Qy	897	TTAATCGG	AAAGACTACTGGCTTTGTGAAAACAGCTGGGGCCCTCAACTTTGGTGAACCAAG	956						
Ib	1122	TTAATCGG	AAAGAAATACTGGCTTTGTGAAAACAGCTGGGGCCCAACTTTGGTGAAGAAG	1181						
Qy	957	GATATATTC	GSATGRCANAAAGATGTGAAATATCACTGTGGATTTGTAGTTATCCCTCTT	1016						
Ib	1182	GATATATTC	CGATGGCAAGAAATAAAGGAATCATTTGTGGATTTGCTAGCTTTCCCTCTT	1241						
Qy	1017	ACCAGAAAT	CTTA---GACCTCTTCATTTTATTAACAGTCCCAAAAATTAAGAACACTTTC	1073						
Ib	1242	ACCAGAAAT	CTTAGAGATCTCTCTTTTATTAACAAATCAGAAATATTAAGACACTTTC	1301						

```

Qy 1074 TCTTAATTTAAATTTTAACTGCTGATA-ATAGTAAAAAATAATGTGTGATGA 1122
      ||||| ||||| ||||| ||||| || ||||| ||||| |||||
Db 1302 TCTTAACTTAATTTTCTGCTGTATCCAGAAGAAATAATTTGTGTCATGA 1351

RESULT 7
US-09-716-990-1545
: Sequence 1545, Application US/09716990
: GENERAL INFORMATION:
: APPLICANT: Gearing, David P.
: APPLICANT: Holtzman, Douglas A.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: TITLE OF INVENTION: THEREFOR
: FILE REFERENCE: 1600.2039-001
: CURRENT APPLICATION NUMBER: US/09/716,990
: CURRENT FILING DATE: 2000-11-21
: PRIOR APPLICATION NUMBER: US 60/166,926
: PRIOR FILING DATE: 1999-11-22
: NUMBER OF SEQ ID NOS: 1691
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1545
: LENGTH: 2205
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)..(2205)
: OTHER INFORMATION: n = A,T,C or G
US-09-716-990-1545

```


Db	762	TGGATTTGCTCAACTGAAAAATATGAAACAAAGCGTGCAGTGGTGGCTTCATGACAAACGG	821
QY	597	CTTTCACGATATATTTGATACAAACAGGCGATTGATTCAGAGAGCTTTCCTATACCTCAAAAG	656
Db	822	CTTTCACGATCATATTGATACAAAGGCGATTCGACTCAGACGTTCCCTATCCCTACAAAG	881
QY	657	CCATGAATGGAAAGTCGACATATAGCTCAAAAAGCGAGCTCCACATGTTCAAAAGTATA	716
Db	882	CCATGGATCAGAAATGTCATATGATCAAAATATCGTCTGCCACATGTTCCAAAGTACA	941
QY	717	CTGAACCTCCCTTTGGCAGTGAAGATGCCCTTAAAAAGAGCTGTGGCCAAATAAAGGACCTG	776
Db	942	CHCAACTTCCTTATGCCAGAGAAGATGCTCTGAAGAAGCTCTGGCCAAATAAAGGCCAG	1001
QY	777	TGCTGTGGCTATAGATGGAGCGACTATCTCTTCTCTGTACAGAAGTGGTGTCTACT	836
Db	1002	TGCTGTGGTGTAGATGGCGGTCACTCTCTCTCTCTACAGAAGTGGTGTCTACT	1061
QY	837	ATGAACCATCCTCTACTCAGAAATGTAATCATCGAGTATTAGTGGTTGGCTATGGTAAACC	896
Db	1062	ATGAACCATCCTGTACTCAGAAATGTAATCATCGAGTATTAGTGGTTGGCTATGGTATC	1121
QY	897	TTAATGGGAAGACTACTGGCTGTGAAAACAGCTGGGGCCTCAACTTTGGTGAACCAAG	956
Db	1122	TTAATGGGAAGAAATACTGGCTGTGAAAACAGCTGGGGCCTCAACTTTGGTGAAGAAAG	1181
QY	957	GATATATTCGGATGGCAAGAAACAGTGGAAATCACTGTGGGATTGCTAGTTATCCCTCTT	1016
Db	1182	GATATATTCGGATGGCAAGAAATCACTGTGGGATTGCTAGTTATCCCTCTT	1241
QY	1017	ACCAGAAATCTA---GACCTCTTCAATTTTATTAACAGTCCCAAAAATTTGAAACACTTTC	1073
Db	1242	ACCAGAAATCTAAGAGATCTCTCTCTTTTATTAACAAATCAAGAAATATGAAGCACATTC	1301
QY	1074	TCCTAATTAATTTTACCTCTCTATA-ATAGTAAAAATAAAATGTGTCATGA	1122
Db	1302	TCCTAATTAATTTTCCCTCTCTATCCAGAAATAATTTGTCATGA	1351
RESULT 8			
US-09-721-589-5798			
; Sequence 5798, Application US/09721589			
; GENERAL INFORMATION:			
; APPLICANT: Gearing, David P.			
; APPLICANT: Villevial, Jean-Luc			
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES			
; TITLE OF INVENTION: THEREFOR			
; FILE REFERENCE: 1600.2045-001			
; CURRENT APPLICATION NUMBER: US/09/721,589			
; CURRENT FILING DATE: 2000-11-22			
; PRIOR APPLICATION NUMBER: 60/167,380			
; PRIOR FILING DATE: 1999-11-24			
; NUMBER OF SEQ ID NOS: 7017			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 5798			
; LENGTH: 2205			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc_feature			
; LOCATION: (1)..(2205)			
; OTHER INFORMATION: n = A,T,C or G			
US-09-721-589-5798			

RESULT 9

US-09-726-172-2249

```
; Sequence 2249, Application US/09/726172
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2052-001
; CURRENT APPLICATION NUMBER: US/09/726,172
; PRIOR FILING DATE: 2000-11-29
; PRIOR FILING DATE: 2000-11-29
; NUMBER OF SEQ ID NOS: 2733
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2249
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (1) (2205)
; OTHER INFORMATION: n - A,T,C or G
US-09-726-172-2249
```

Query Match 76.1%; Score 856.4; DB 29; Length 2205;
Best Local Similarity 87.3%; Pred. No. 1.8e-226;
Matches 986; Conservative 0; Mismatches 136; Indels 8; Gaps 4;

```
QY 1 GCATTATTAGTATGGAGCACTGAGGTTCTTTC---ATGAAATGGCTAGTTGGGTCG 57
DB 222 GGACTCTTACTGTGGAGCACTGCTGCTTCTATCACAATGAACGGCTGTTGTGTGC 281
QY 58 TTCCTGTGCTCCTATGAGTGGPACAAAGTACATAAAGATCCCACTCTGGATCATCACT 117
DB 282 TCTTGTGTGCTCCTGTCAGTGGCAGCTTGCATATAAGATCTCCCTGGATCACCCT 341
QY 118 GGAATCTCGGAAGAAACCTACAGCAACAAT-CAAGAGAGAGAAATGAGGATGAC 176
DB 342 GGATCTCGGAAGAAACCTATGCGCAACAATACAGAAAGAAATGAAGACAGTAC 401
QY 177 GGGGTCTCATCTGGAAGAAATCTAAATTTGTGATGCTTCCAACTCTGGAACACTCA 236
DB 402 GAGCTCTCATCTGGAAGAAATCTAAAGTTTGTGATGCTTCCAACTCTGGAACACTCA 461
QY 237 TGGAAATGCACTATGATGATGAGTATGAAATAT-TGGAGAAATGA-TGTGTGAAGAG 296
DB 462 TGGAAATGCACTATGATGATGAGTATGAAATAT-TGGAGAAATGA-TGTGTGAAGAG 521
QY 297 TGATATCTTGAATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 356
DB 522 TGATATCTTGAATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 581
QY 357 GGTCAAACTCTAATCAAAATTCGTTATTTTGTGAGTATGAGTATGAGTATGAGTATG 416
DB 582 AGTCAAACTCTAATCAAAATTCGTTATTTTGTGAGTATGAGTATGAGTATGAGTATG 641
QY 417 CTGAAATGCACTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 476
DB 642 CTGAAATGCACTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 701
QY 477 TGAAGCACTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 536
DB 702 TGAAGCACTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 761
QY 537 TGATGCTCACTGAAATATGGAATTAAGGCTGCAATGAGGCTTCCATCCCTACAAG 596
DB 762 TGATGCTCACTGAAATATGGAATTAAGGCTGCAATGAGGCTTCCATCCCTACAAG 821
QY 597 CTTTCAGTATATTATGATACAGGGCATTTGATTCAGAGCTTCCATCCCTACAAG 656
DB 822 CTTTCAGTATATTATGATACAGGGCATTTGATTCAGAGCTTCCATCCCTACAAG 881
QY 657 CCATGAATGGAAGTCCAGATATGACTCAAAAAAGGAGCTGCCACATGTTTCAAGATATA 716
```

```
DB 882 CCAATGATCAAAAATGCAAAATACAAATACAAATACAAATACAAATACAAATACAA 941
QY 717 CTGAATTTCTGCTTGTGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 776
DB 942 CTGAATTTCTGCTTGTGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 1001
QY 777 TGTCTGTGCTATAGATGCGAGGCACTATTTCTTCTTCTTCTTCTTCTTCTTCTTCT 836
DB 1002 TGTCTGTGCTATAGATGCGAGGCACTATTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1061
QY 837 ATGAACCACTGCTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 896
DB 1062 ATGAACCACTGCTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 1121
QY 897 TTAATGGAAGACTACTGCTTGTGAAACAGCTGGAACAGCTGGAACAGCTGGAACAG 956
DB 1122 TTAATGGAAGACTACTGCTTGTGAAACAGCTGGAACAGCTGGAACAGCTGGAACAG 1181
QY 957 GATATATTGCGATGCGAGCAAGCACTGGAACAGCTGGAACAGCTGGAACAGCTGGA 1016
DB 1182 GATATATTGCGATGCGAGCAAGCACTGGAACAGCTGGAACAGCTGGAACAGCTGGA 1241
QY 1017 ACCAGAAATCTA---GACCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1073
DB 1242 ACCAGAAATCTACAGGATCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1301
QY 1074 TCTTAATTAATTTTACTGCTCTATA-ATAGTAAATTAATTAATTAATTAATTAAT 1122
DB 1302 TCTTAATTAATTTTACTGCTCTATA-ATAGTAAATTAATTAATTAATTAATTAAT 1451
```

RESULT 10

```
US-09-726-788-6295
; Sequence 6295, Application US/09726788
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Kingsbury, Gillian A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2051-001
; CURRENT APPLICATION NUMBER: US/09/726,788
; CURRENT FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 7691
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6295
; LENGTH: 2205
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; LOCATION: (1) (2205)
; OTHER INFORMATION: n - A,T,C or G
US-09-726-788-6295
```

Query Match 76.1%; Score 856.4; DB 29; Length 2205;
Best Local Similarity 87.3%; Pred. No. 1.8e-226;
Matches 986; Conservative 0; Mismatches 136; Indels 8; Gaps 4;

```
QY 1 GCATTATTAGTATGGAGCACTGAGGTTCTTTC---ATGAAATGGCTAGTTGGGTCG 57
DB 222 GGAATCTTAT-TGTGGAAGCAAA-TG-TGTGTTCTAT-A-ATTAAGAAAGG-TGTGTTGTG 281
QY 58 TTCCTGTGCTCCTATGAGTGGCAGCACTACATAAAGATCCCACTCTGGATCATCACT 117
DB 282 TCTTGTGTGCTCCTGTCAGTGGCAGCTTGCATATAAGATCTCCCTGGATCACCCT 341
QY 118 GGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGAGAGAGAAATGAGGATGAC 176
DB 342 GGCATCTCTGGAAGAAACCTATGGCAACAATACAAAGTCAAAATTAATTAATTAATTAAT 401
```


NUMBER OF SEQ ID NOS: 5415
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 5142
LENGTH: 2205
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(2205)
OTHER INFORMATION: n = A,T,C or G
US-09-726-811-5142

Query Match 76.1%; Score 856.4; DB 29; Length 2205;
Best Local Similarity 87.3%; Pred. No. 1.8e-226;
Matches 986; Conservative 0; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATTAGTATGGGAGCACCTCGAGGTTCTTTTC---ATGAATGGCTAGTTGGGCTGC 57
DB 222 GGACTTTACTGTGGAGCACTGCTGGTTCTATCACAATGAAGCGGCTGTTGTGTC 281

QY 58 TTCCGTTGCTCTATGAGTGGGACAAAGTACATAAGATCCCACTCTGGATCATCACT 117
DB 282 TCTTGGTGTCTCTCGAGTGGCACAGTTCATCAAAAGATCCTACCCCTGGATCACCAC 341

QY 118 GGAATCTCTGGAAGAAACCTACAGTAAACAAT-CAAGGAAGAGATGAGGAAGTAGCAC 176
DB 342 GGCATCTCTGGAAGAAACCTATGGCAAACAATACAGGAAGAAATCAAGAACGATAC 401

QY 177 GCGCTCTCATCTGGGAAAAAATCTAAAAATTTGTATGCTTCACAACTCTGGAACTCAA 236
DB 402 GACGCTCATCTGGGAAAGAAATCTAAAGTTTGTATGCTTCACAACCTGGAGCAATCAA 461

QY 237 TGGGAATGCATTCATATGATCTAGGCATGAACCATCTGGAGACATGCTGCTGAAGAAG 296
DB 462 TGGCAATGCATTCATATGATCTAGGCATGAACCATCTGGAGACATGCTGCTGAAGAAG 521

QY 297 TGATATCTTTGATGGGTTCCCTGAGAGTTCCCAAGCAATGGCAGAGAAATGTCATTATA 356
DB 522 TGATCTCTTTGATGGGTTCCCTGAGAGTTCCCAAGCAATGGCAGAGAAATGTCATTATA 581

QY 357 GGTCAAACTTAATCAAAATTCCTGATCTGAGCTGGAGTGGAGAGAGAGGCTGTGTTA 416
DB 582 AGTCAAACTTAATCGGATATGCTGATCTGAGCTGGAGTGGAGAGAGAGGCTGTGTTA 641

QY 417 CTGAAGTCAAAATCCAGGGTTCTTTGGTCTCTTTGGGCTTTTCAGGCTGTGGGGGCC 476
DB 642 CTGAAGTCAAAATCAAGGTTCTTTGGTCTCTTTGGGCTTTTCAGGCTGTGGGGGCC 701

QY 477 TGGAGACAACTGAAGTAAACAGAGAGAGGCTGCTCTGAGTGGACAGAACTGG 536
DB 702 TGGAGACAACTGAAGTAAACAGAGAGAGGCTGCTCTGAGTGGACAGAACTGG 761

QY 537 TAGATTGCTCAACTCAAAATATAGGAATAAAGGCTGCAATGGCGTTTCATGACAACTG 596
DB 762 TGAATTGCTCAACTCAAAATATAGGAATAAAGGCTGCAATGGCGTTTCATGACAACTG 821

QY 597 CTTTCCAGTATATTATGATTAACAGGCAATTTGATCAGAGCTTCTATCCCTACAAAG 656
DB 822 CTTTCCAGTATATTATGATTAACAGGCAATTTGATCAGAGCTTCTATCCCTACAAAG 881

QY 657 CCATGAATGGAAATGTCAGATATGACTCAAAAGAGGAGCTGCCACATGTTCAAGTATA 716
DB 882 CCATGAATGCAAAATGTCAGATATGACTCAAAAGAGGAGCTGCCACATGTTCAAGTATA 941

QY 717 CTGAATTCCTCTTGGAGTGAAGTGCCTTTAAAGAGAGCTGTGCCCAATTAAGGACCTG 776
DB 942 CTGAATTCCTCTTGGAGTGAAGTGCCTTTAAAGAGAGCTGTGCCCAATTAAGGACCTG 1001

QY 777 TGTCTGTGCTATAGATCGAGCCACTATCTTTCTCTCTACAGAAAGTGGTGTCTACT 836
DB 1002 TGTCTGTGCTATAGATCGAGCCACTATCTTTCTCTCTCTACAGAAAGTGGTGTCTACT 1061

QY 837 ATGAACCACTCTCTACTCAGAAATGTAATCATGAGAGTATTAGTGGTGGCTATGGTAACC 896

DB 1062 ATGAACCACTCTCTACTCAGAAATGTAATCATGTAAGTGTGAGTGTGATGTCATC 1121
QY 897 TTAATGGGAAGACTACTGGCTTTGTGAAAAACAGCTGGGCTCAACTTTGGTGACCAAG 956
DB 1122 TTAATGGGAAGACTACTGGCTTTGTGAAAAACAGCTGGGCTCAACTTTGGTGAGGAAG 1181
QY 957 GATATATTGAGTGGCAAGAAACAGTGGAAATCAATTTGGGATTTGTTATCTCTCTT 1016
DB 1182 GATATATTGAGTGGCAAGAAATGAAGAAATCAATTTGGGATTTGTTATCTCTCTT 1241
QY 1017 ACCAGAAATCTA---GACCTCTTCATTTTATACAGTCCAAAAAATTTGAAATCTTTC 1073
DB 1242 ACCAGAAATCTAGAGGATCTCTCTCTTTTATAAATAATCAAGAAATTAAGACATTTTC 1301
QY 1074 TCTTAATTTAATTTACCTGCTATA-ATAGTAAAAAATAATGTCATCA 1122
DB 1302 TCTTAATTTAATTTTCTGCTGTATCCAGAAATAATTTGTGTCATCA 1351

RESULT 14

US-09-760-475-1251
Sequence 1251, Application US/09760475

GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT249
CURRENT APPLICATION NUMBER: US/09760,475
CURRENT FILING DATE: 2001-01-16
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 4122
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1251
LENGTH: 2325
TYPE: DNA
ORGANISM: Homo sapiens
US-09-760-475-1251

Query Match 76.0%; Score 856; DB 30; Length 2325;
Best Local Similarity 87.2%; Pred. No. 2.4e-226;
Matches 985; Conservative 1; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATTAGTATGGGAGCACCTCGAGGTTCTTTTC---ATGAAATGGCTAGTTGGGCTGC 57
DB 86 GCACTCTTACTGTGGAGCACTGCTGGTTCTATCACAATGAAACGGCTGTTGTGTC 145

QY 58 TTCCGTTGCTCTCTATGAGTGGCACAAGTACATAAGATCCCACTCTGGATCATCACT 117
DB 146 TCTTGGTGTCTCTCTGCACTGGCAGTTGCAATAAGATCCTACCTTGTGATACCACT 205

QY 118 GGAATCTCTGGAAGAAACCTTACAGCAAAACAAT-CAAGGAAGAGAAATGAGGAAGTAGCAC 176
DB 206 GGCATCTCTGGAAGAAACCTTATGGCAAAACAATACAGGAAGAAATGAGGAAGTAGCAC 265

QY 177 GCGCTCTCATCTGGGAAAAAATCTAAAAATTTGTGATGCTTCACAACTCTGGAACCTCAA 236
DB 266 GACGCTCTCATCTGGGAAAGAAATCTAAAGTTTGTGATGCTTCACAACTCGAGCATCAA 325

QY 237 TGGGAATGCATTCATATGATCTAGGCATGAACCACTCTGGGAGACATGACTGGTCAAGAG 296
DB 326 TGGGAATGCATTCATATGATCTGGCATGAACCACTCTGGGAGACATGACCAGTAGGAAG 385

QY 297 TGATATCTTTGATGGGTTCCCTGAGAGTTCCCAAGCAATGGCAGAGAAATCTCACTTATA 356
DB 386 TGATGCTTTGATGAGTTCCCTGAGAGTTCCCAAGCAATGGCAGAGAAATATCACATATA 445

QY 357 GGTCAAACTCTAATCAGAAATTTGCTGATTTCTGGACTGGAGAGAGAGGTTGTGTTA 416
DB 446 AGTCAAACTCTAATCCGATATTTGCTGATTTCTGGACTGGAGAGAGAGGTTGTGTTA 505

QY 417 CTGAAGTCAAAATCCAGGCTTCTTGTGCTGTGTTGGCTTTCAGGCTCTCTGGGGGCC 476
DB 506 CTGAAGTCAAAATCCAGGCTTCTTGTGCTGTGTTGGCTTTCAGGCTCTCTGGGGGCC 565

Qy	477	TGGAAGCACAACCTGAAGCTAAAACACGAAAGCTGGCTCTCTGAGTGCACAGAACCTGG	536
Db	566	TGGAAGCACAGCTGAAGCTCAAAACAGSAAATGTGTGTCTCTAGTGGCCAGCAACCTGG	625
Qy	537	TAGATTCCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCGTTTCATCACAACCTG	596
Db	626	TGGATTCCTCAACTGAAAAATATGGAACAAGGCTGCAATGGTGGCTTCATCACAACGG	685
Qy	597	CTTTCCAGTATATTTTGATACGAAGGGCAATGATTACAGAAGCTTGCTATCCCTACAAAG	656
Db	686	CTTTCCAGTACATCATTTGATTAACAAGGGCATCCAGCTCAGACGGTTGCTATCCCTACAAG	745
Qy	657	CCATGAATGGAAGTGCAGATATCACTTCAAAAAAGGCAGCTGCCACATGTTCAAAAGTATA	716
Db	746	CCATGGATCAGAAATGTCAATATGACTCAAAATATCGTGTGCCACATGTTCAAGGTACA	805
Qy	717	CTGAACCTCCCTTTGGCAGTGAAGATGCCCTTAAAGAAGCTGTGGCCAATAAAGGACCTG	776
Db	806	CTGAACCTCCCTATATGGCAGAGAAGATGCTCTGAAGAAGCTGTGGCCAATAAAGGCCAG	865
Qy	777	TGCTGTGGGTATAGATGTGAGCGACTATTCCTTTCTTCCTGTACAGAAGTGGTGTCTACT	836
Db	866	TGCTGTCTGTGTAGATGCGCGTCATCCTCTCTTTCTCTCCTCTACAGAAGTGGTGTCTACT	925
Qy	837	ATCAACCATCCTGTACTACAATGTCAATCATCGAGTATTAGTGGTTGGCTATGTTGAACC	896
Db	926	ATGACCATCCTGTACTCAGAAATGTAAATCATGGTGTACTGTGTGGTATGTTGTGATC	985
Qy	897	TTAATGGGAAGACTACTGGCTTGTGAAAACAGCTTGGGGCCTCAACTTTGGTGGACCAAG	956
Db	986	TTAATGGGAAGAAATACTGGCTTGTGAAAACAGCTTGGGGCCACAACCTTTGGTGAAGAAG	1045
Qy	957	GATATATTCGGATGGCAAGAAACAGTGGAAATCACTGTGGGATGTGAGTTATCCCTCTT	1016
Db	1046	GATATATTCGGATGGCAAGAAATAAGAGAAATCATTTGGGATTTGCTAGCTTTCCCTCTT	1105
Qy	1017	ACCAGAAATCTA--GACCTCTTCATTTTATACAAAGTCCAAAAATTTGAAACACTTTC	1073
Db	1106	ACCAGAAATCTAGAGATCTCTCCTTTTATTAACAATTCAGAAATATGAGCACTTTC	1165
Qy	1074	TCCTTAATTTAAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA	1122
Db	1166	TCCTTAACCTAAATTTTCTGCTGCTATCCAGAGAAATAATTTGTGTGTCATGA	1215

RESULT 15

```

US-09-760-475-25
; Sequence 25, Application US/09760475
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ49
; CURRENT APPLICATION NUMBRK: US/09760.475
; CURRENT FILING DATE: 2001-01-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4122
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2350
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-760-475-25

```

	Query Match	76 %	Score 856	DB 30	Length 2350
	Best local Similarity	87 %			
	Matches 985	Conservative 1	Mismatches 136	Indels 8	Gaps 4
QY	1	GCATTATTAGTATGGGAGCACTGCAGTGTCTTTC	---	ATGAATGGCTAGTGTGGCTGC	57
Db	111	GGACTCTTACTGTGGAGCAACTGCTGGTCTATCAATGAAACGGCTGTGTGTGC			170
QY	58	TTCCGTTGTGCTCTCTATGCAGTGGCAGTACATAAAGATCCCACTCTGGATCATCACT			117

Db	171	TCTTGGTGTGCTTCTCGAGTGGCACAGTTGCATAAAGATGCTTACCTCTGGATCA/CAC/T	233	
QY	118	GGAATCTCTGGAA/AAAAACCTACACGAACAACAT	CAGGGAAGAGGAATAGAGAAATAGAGAC	176
Db	231	GGCATCTCTGGAAGAAACCTATGCGAAACAATACAGGAAAGAAATCAAGGAAGGAGTAC	290	
QY	177	GGCGTCTCATCTGGGGAAAAAATCTAAAATTTTGTGATGCTTCA/AA/TGTGTAA/AC/TAA	246	
Db	291	GACGTCTCATCTGGGAAAGAATCTAAAAGTTTGTGATGCTTCA/CAACCTCTGGAGCATTCAA	350	
QY	237	TGGGAATGCATTATATGATCTAGGCATCAACCATCTGGGAGACATGATCTGCTGAAGAAG	296	
Db	351	TGGGAATGCATCATACCATCTGGGCATGAACCACTGGGAGACATGACCACTGGAAGAAG	410	
QY	297	TGATATCTTTTCATGGTTCTCTGAGTCTCTCA/AA/TAATGTAA/AA/AAATTTCACTTATA	356	
Db	411	TCATGCTCTTATGATGAGTTCCCTGAGAGTTCCGAGCAGTCCGAGCAGTGGCAGAGAAATATCAGATATA	470	
QY	357	GCTCAAACTCTAATCAGAAATTCGCTGATCTCTGCACCTGGCAGACAGAAGTCTGTGTTA	416	
Db	471	AGTCAAACTCTAATCGGATATTTCT/ATTTTGTAA/TGAGAGAGAAAGGTGTGTTA	530	
QY	417	CTGAAGTGAATACACAGGGTTCTTGTGTGCTTGTGTGGCTTTTACGGCTCTGAGGAGCC	476	
Db	531	CTGAAGTGAATATCAAGGTTCTTGTGTGCTTGTGTGGCTTTTACGGCTCTGAGGAGCC	590	
QY	477	TGGAGCAACAAGTGAAGCTAAACACAGCAAAAGCTGTCTCTGATGTGCACAGAA/TGG	536	
Db	591	TGGAGCACAGCTTAAGCTGAACAGAGGAAGCTGGTGTCTCT/AGTGGCT/AGAACTTGG	650	
QY	537	TAGATTGCTCAACTCAAAATATGGAATTAAGGCTGCAATGGGTCTTTCATACAACTG	596	
Db	651	TGGATTGCTCAACTGAAATATATGAAC/AAAGG/TGC/AA/TGT/AGCTTCATGAA/ACGG	710	
QY	597	CTTTCCAGTATATTATCATACCAAGGCGATTGATTCAGAAGCTTCCTAGTGTCA/AAAG	656	
Db	711	CTTTCCAGTACATCATTCATACCAAGGCGATCGACTCAGCGCTTCCATATCTTACAAG	770	
QY	657	CCATGAATGAAATGTAGATATATCT/AAAAA/AT/ATGCAACATGTT/CAAGTATA	716	
Db	771	CCATGGATCAGAAATGTCAATATGACTCAAAATATGCTGCTGCCACATGTT/CAAGTACA	830	
QY	717	CTGAACCTCCCTTTGGCAGTCAAGATGCTTTAAAAGAAGCTCTGCCAATAAAGCACTGG	776	
Db	831	CTGAACCTCTTATGGCAGAGAGATGTCTGTAAAGAAGCTGTGTGCATTAAGGCTCCAG	890	
QY	777	TGCTGTGGCTATAGATCGCAGCCACTATCTTTCTTCTCTGATACAGAAGTGTGCTTACT	836	
Db	891	TGCTGTGTGTAGATCGCGCTCATCTCTTCTTCTCTCTATACAGAAGTGTGCTTACT	950	
QY	837	ATGAACCATCTGTACTCAGAAATGTGAATCATGAGTATTAGTGGTGGCTATGGTAACC	896	
Db	951	ATGAACCATCTGTACTCAGAAATGTGAATCATGAGTATTAGTGGTGGCTATGGTAACC	1010	
QY	897	TTAATGGGAAAGACTACTGGCTGTGTGAAGAAACAGCTGGGGCTCAACTTTTGTGACAC	956	
Db	1011	TTAATGGGAAAGAACTACTGGCTGTGTGAAGAA/AA/TGGG/AA/AA/AACTTTGTGGAAG	1070	
QY	957	GATATATTCGATGGCAGAAACACTGTGAAAATCACTGTGGGATTTGCTACTTATGGCTTT	1014	
Db	1071	GATATATTCGATGGCAGAAATAAAGGAAACATCACTGGGATTTGCTACTTATGGCTTT	1130	
QY	1017	ACCCAGAAATCTA---GAACTTCTTCAATTATTA/AA/AACTT/AAAAATTTGAACACTTTC	1077	
Db	1131	ACCCAGAAATCTACAGGAATCTCTCTCTTTTATACAAATCAAGAAATATGAAGTACTTTC	1190	
QY	1074	TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTGATAGA	1122	
Db	1191	TCTTAAGTAAATTTTCTGCTGATATGACAGAAAGAAATTAATTTGTCTATGA	1240	

us-10-010-580-1.rnrm

Tue Jul 1 14:40:55 2003

Job time : 2968 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 09:28:21 ; Search time 978 Seconds

(without alignments)
8229.219 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1126

Sequence: 1 gcatattagtagggagca.....ataaaatgtgtcatgacct 1126

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched 8166257 seqs. 3573790582 residues

Total number of hits satisfying chosen parameters: 1633514

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA, New: *

- 1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq2:*
- 7: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq3:*
- 8: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
- 9: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq2:*
- 10: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq3:*
- 11: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*
- 12: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq2:*
- 13: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	ID	Description
1	856	76.0	2325	8	US-10-227-425-1251
2	856	76.0	2350	8	US-10-227-425-25
3	854.8	75.9	1514	10	US-10-170-235-26187
4	853.6	75.8	1514	12	US-60-452-680-10960
5	853.6	75.8	1514	12	US-60-453-135-6918
6	853.6	75.8	1514	12	US-60-453-050-6918
7	853.6	75.8	1514	13	US-60-455-444-3608
8	853.6	75.8	1514	13	US-60-465-241-3608
9	853.6	75.8	1514	13	US-60-466-412-6918
10	845.4	75.1	1763	5	US-09-873-367C-173
11	845.4	75.1	1763	10	US-10-219-051B-2756
12	845.4	75.1	1763	10	US-10-219-051B-10848
13	842.4	74.8	1255	6	US-10-342-887-987
14	824.8	73.3	1366	6	US-09-724-676-33122
15	824.8	73.3	1366	6	US-09-724-676A-33122
16	824.8	73.3	1409	6	US-09-724-676-33121
17	824.8	73.3	1409	6	US-09-724-676A-33121
18	709.2	63.0	1895	1	PCT-US03-09774-96
19	665.8	59.1	1290	9	US-10-144-771-6528
20	634.2	56.3	1347	10	US-10-256-911-157

c

ALIGNMENTS

RESULT 1

US-10-227-425-1251
Sequence 1251, Application US/10227425
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT249CIN
CURRENT APPLICATION NUMBER: US/10/227,425
CURRENT FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 09/760,475
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/218,290
PRIOR FILING DATE: 2000-07-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 4122
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1251
LENGTH: 2325
TYPE: DNA
ORGANISM: Homo sapiens
US-10-227-425-1251

Query Match 76.0%, Score 856, DB 8, Length 2325;
Best Local Similarity 87.2%, Pred. No. 14e+20,
Matches 985; Conservative 1; Mismatches 136; Indels 8; Gaps 4;

QY

1 GCATTATTAGTATGGGACACCTCGAGTCTTTC---ATCAATGCTAGTGGGCGC 57

Sequence 156, App
Sequence 3409, App
Sequence 2945, App
Sequence 1700, App
Sequence 949, App
Sequence 2002, App
Sequence 155, App
Sequence 3409, App
Sequence 2754, App
Sequence 10846, A
Sequence 2945, App
Sequence 831, App
Sequence 8754, App
Sequence 243045,
Sequence 1, Appl1
Sequence 1556, App
Sequence 479, App
Sequence 6305, App
Sequence 11530, A
Sequence 200, App
Sequence 26186, A
Sequence 10959, A
Sequence 6917, App
Sequence 6917, App
Sequence 3607, App

634.2 56.3 1348 10 US-10-256-911-156
631 56.0 1330 1 PCT-US03-03482-3409
631 56.0 1330 1 PCT-US03-03194-2945
631 56.0 1330 9 US-10-152-319A-1700
631 56.0 1330 9 US-10-301-856-949
631 56.0 1330 9 US-10-338-044-2002
631 56.0 1330 10 US-10-256-911-155
631 56.0 1330 10 US-10-357-507-3409
631 56.0 1330 10 US-10-219-051B-2754
631 56.0 1330 10 US-10-219-051B-10846
631 56.0 1330 12 US-60-436-643-2945
631 56.0 1330 13 US-60-468-720-831
631 56.0 671 1 PCT-US03-13853-8754
332 6 29.5 520 6 US-09-912-293-243045
310 27.5 1482 10 US-10-318-584-1
310 27.5 1669 9 US-10-240-425-1556
310 27.5 1669 9 US-10-342-887-479
310 27.5 1669 10 US-10-219-051B-6305
310 27.5 1669 10 US-10-219-051B-11530
310 27.5 1702 11 US-60-474-733-200
310 27.5 1702 10 US-10-170-235-26186
308.4 27.4 1942 12 US-60-452-680-10959
306.8 27.2 1942 12 US-60-453-135-6917
306.8 27.2 1942 12 US-60-453-050-6917
306.8 27.2 1942 13 US-60-455-444-3607

Db	86	GGACTCTTACTGTRGGAGCAACTGCTGGTTCTCTATCAAAATGAAACGGCTGGTTGTGTGC	145
Qy	58	TTCCGGTTGTCTCCTATGTCAGTGGCAACAAGTACATAAAGATCCACACTCTGGATCATCACT	117
Db	146	TCTTGGTGTGCTCCTCTGTGCAGTGGCAGATTGCATAAAGATCTTACCTGGATCACCACT	205
Qy	118	GGAACTCTCGAAGAAAACCTACAGCAACAAT--CAAGGAAGAGAAATGAGGAAGTAGCAC	176
Db	206	GGCATCTCTGGAAAGAAAACCTATGGCAACAAATACAAAGGAAAAAGATGAAGAAGCAGTAC	265
Qy	177	GGCGTCTCATCTGGGAAAAAAATCTAAAATTTGTGATGCTTCAAAATCTCGAACACTCAA	236
Db	266	GAGCTCTCATCTGGGAAAAAGAAATCTAAAGTTTGTGATGCTTCAAACTCGAGCATTCAA	325
Qy	237	TGSGAAATGCAATTCATATGATCTAGGCATGAACCATCTGGAGACATGACTGGTGAAGAAG	296
Db	326	TGGGAATGCACTCATACGATCTGGGATGAACCACTGGCAGACATGACCAAGTCAACAAG	385
Qy	297	TGATATCTTTGATGGTTCCCTCAGAGTTCCCGAGCCAAATGGCAGAGAAAATGTCACATTATA	356
Db	386	TGATGCTTTTGTATGAGTTCCCTCAGAGTTCCCGAGCCAAATGGCAGAGAAAATGTCACATTATA	445
Qy	357	GGTCAAACTCTAATCAGAAAATCCCTGATTTCTGTGGACTGGAGAGAGAAGGGTGTGTTA	416
Db	446	AGTCAAAACCTTAATCGATATTGCCATGCTGTGGACTGGAGAGAAAAGGGTGTGTTA	505
Qy	417	CTCAAGTGAATACCAAGGGTTCTGTGGTGCCTTGTGGGCTTTTCAGGCGCTGTGGGGCCC	476
Db	506	CTCAAGTGAATATCAAGGGTTCTGTGGTGCCTTGTGGGCTTTTCAGTGCCTGTGGGGCCC	565
Qy	477	TGGAAGCACAACTCAAGCTAAAAACAGGAAGCTGGTGTCTCTGAGTGCACAGAACTTGG	536
Db	566	TGGAAGCACAGCTGAAGCTGAAAACAGGAAGCTGGTGTCTCTGAGTGCACAGAACTTGG	625
Qy	537	TGATTTGCTCAACTGAAAATATGGGAATAAAGGCTGCAATGGCGGTTTCATGACAACTG	596
Db	626	TGATTTGCTCAACTGAAAATATGGGAATAAAGGCTGCAATGGCGGTTTCATGACAACTG	685
Qy	597	CTTTCCAGTATATATTATTGATAAATAGGGCAATTGATTCAAGATTCTCTATCCCTACAAAG	656
Db	686	CTTTCCAGTATACATTTGATAACAGGGCATCGACTCAGAGCTTCTATCCCTACAAAG	745
Qy	657	CCATGAATGAAAGTCCAGATATGACTCAAAAAGCGAGCTGCCACATGTTCAAAAGTATA	716
Db	746	CCATGATCAGAAATGTCAATATGATCAAAATATGTGCTGCCACAATGTCAAAATACA	805
Qy	717	CTCAACTTCCTTTTGGCAGTGAAGATGCCCTTAAAGAAAGCTGTGGCAATAAAGGACCTG	776
Db	806	CTCAACTTCCTTTATGGCAGAGAGATGTCTGTGAACAGAGCTGTGGCAATAAAGGCCAG	865
Qy	777	TGCTGTGGGCTATAGATGCGAGCACTATTCTTTCTTCTGTATAGAAAGTGGTGTCTACT	836
Db	866	TGCTGTGGTGTAGATGGCGGTATCCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	925
Qy	837	ATCAACCATCTCTACTCAGAATGTGAATCATGGAGTATTTAGTGGTTGGCTATGTTAACC	896
Db	926	ATGAACCATCTCTACTCAGAATGTGAATCATGGTGTACTTGTGGTGGCTATGTTGATC	985
Qy	897	TTAATGGGAAGACTACTGGCTTTCTGAAAAACAGCTGGGGCCCTCAACTTTGGTGCACCAAG	956
Db	986	TTAATGGGAAGAAATACTGGCTTTGTGAAAAACAGCTGGGGCCCACTTTGGTGAAGAAG	1045
Qy	957	GATATATTGATGAGCAAAAGAGTGGAAATCAATCTGTGGATTGTAGTTATCCCTCTT	1016
Db	1046	GATATATTGCGATGGCAAGAAATAAAGGAAATCAATCTGCGATTTGTAGCTTTCCCTCTT	1105
Qy	1017	ACCCAGAAATCTA---GACCTCTTCAATTTTATAACAGTCCCAAAAATTCAAACACTTTC	1073
Db	1106	ACCCAGAAATCTAGAGATCTCTCCCTTTTATACAAATCAGAAATATGAGACACTTTC	1165
Qy	1074	TCCTTAATTTAAATTTACCTGCTATA-ATAGTAAAAATAAATGTGTCTATGA	1122

```

Db      1166  TCTTAACTTAAATTTTCTGCTGTATCCAGAGAAATAATTCTGTCATGA 1215

RESULT 2
US-10-227-425-25
; Sequence 25, Application US/10227425
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT49CIN
; CURRENT APPLICATION NUMBER: US/10/227,425
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 09/760,475
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
; Remaining prior Application data removed - See File Wrapper or PALL
; NUMBER OF SEQ ID NOS: 4122
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2350
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-227-425-25

```



```

RESULT 4
US-60-452-680-10960
: Sequence 10960, Application US/60452680
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: GROUPE, Andrew
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
: CURRENT APPLICATION NUMBER: US/60/452,680
: NUMBER OF SEQ ID NOS: 116213
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10960
: LENGTH: 1514
: TYPE: DNA
: ORGANISM: Homo sapiens
US-60-452-680-10960

Query Match      75.8%; Score 853.6; DB 12; Length 1514;
Best Local Similarity 86.6%; Pred No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATAGTATGGGAGCACCTGCGAGTTCTTTC---ATGAATGGCTAGTTGGGCTGC 57
DB 224 GGACTCTTACTCTGGGAGCACTGCTGTCTATCACAATGAAGCGCTGTTGTGCGC 283

QY 58 TTCCTGTGCTCTATGCGAGTGGCACAAAGTACATAAGATCCCACTCTGGATCATCACT 117
DB 284 TCTTGTGCTCTGCTGCGAGTGGCACATGCTGATTAAGATCCTACCTGGATCACT 343

QY 118 GGAATCTCTGGAGAAACCTACAGCAACAAT-CAAGGAAGAGAAATGAGGAAGTAGCAC 176
DB 344 GGCACTCTGGAGAAACCTATGGCAACAATCAAGGAAAGAAATGAAGAAGCAGTAC 403

QY 177 GFGCTCTATCTGGGAAACCTATTAATTTGTGATGCTTCACATCTGGACACTCAA 236
DB 404 GAGCTCTATCTGGGAAACCTATTAATTTGTGATGCTTCACACCTGGAGCACTCAA 463

QY 237 TGGGAATGCAATCATATGATCTAGGCATGAACCATCTGGGAGACATGCTGGTGAAGAG 296
DB 464 TGGGAATGCAATCATATGATCTAGGCATGAACCATCTGGGAGACATGCTGGTGAAGAG 523

QY 297 TGATATCTTTGATGGTTCCTTCAGAGTTCCTCAGAGCAATGGCAGAGAAATGTCATTATA 356
DB 524 TGATGCTTTGATGGTTCCTTCAGAGTTCCTCAGAGCAATGGCAGAGAAATGTCATTATA 583

QY 357 GGTCAACTCTAATCAGAAATTCCTGATCTGTGGACTGGAGAGAGAGGCTCTGTATA 416
DB 584 AGTCAAACTCTAATCAGAAATTCCTGATCTGTGGACTGGAGAGAGAGGCTCTGTATA 643

QY 417 CTGAAGTGAATACAGAGTTCCTTCAGAGTTCCTCAGAGCAATGGCAGAGAAATGTCATTATA 476
DB 644 CTGAAGTGAATACAGAGTTCCTTCAGAGTTCCTCAGAGCAATGGCAGAGAAATGTCATTATA 703

QY 477 TGGAGCAACTGGAAGCTAAAACAGGAAAGCTGGTGTCTGTAGTCAGACAGAACTGG 536
DB 704 TGGAGCAACTGGAAGCTAAAACAGGAAAGCTGGTGTCTGTAGTCAGACAGAACTGG 763

QY 537 TAGATGCTCAACTGAAAATATGGAATATAAGGCTGCAATGGCGGTTTCATGACAAGT 596
DB 764 TAGATGCTCAACTGAAAATATGGAATATAAGGCTGCAATGGCGGTTTCATGACAAGT 823

QY 597 CTTTCAGTATATTATGATACAGAGGCAATGATTCAGAGCTTCCTATCCCTACAAG 656
DB 824 CTTTCAGTATATTATGATACAGAGGCAATGATTCAGAGCTTCCTATCCCTACAAG 883

QY 657 CCATGAATGGAAGTGCAGATATGACTCAAAAAGGAGCTGCCACATGTTCAAAAGTATA 716
DB 884 CCATGAATGGAAGTGCAGATATGACTCAAAAAGGAGCTGCCACATGTTCAAAAGTATA 943

QY 717 CTGAATCTCCCTTTGGAGTGAAGATGCTTTAAAAGAAAGCTGTGTGCCCAATAAGAGGCTGC 776

```

```

DB 944 CTGAATCTCCCTATGSSAGAAAGATGCTGGAAGAGCTGTGGCAATAAAGAGCCAG 1004
QY 777 TGTCTGTGGCTATAGATCCGAGCCACTATTTCTTCTCTGTACAGAGTGTGTCTACT 846
DB 1004 TGTCTGTGGTGTAGATGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1063
QY 837 ATGACCATCTCTACTTAGAATGTGAATATAGAGATATAGTGTGTGTGTGTGTGTGT 896
DB 1064 ATGACCATCTCTACTTAGAATGTGAATATAGAGATATAGTGTGTGTGTGTGTGTGT 1123
QY 897 TTAATGGGAAAGACTTACTGCTTGTGAAAACAGCTGGAGCTCAATTTTGGTGCACCAAG 956
DB 1124 TTAATGGGAAAGAAATATGGCTTTTAAAGAAATATGGCTTTTGGTGCACCAAG 1183
QY 957 CATATATTCGGATGGCAAGAACAGTGGAAATCACTGTGGGATTTGGTGTGTGTGTGT 1016
DB 1184 CATATATTCGGATGGCAAGAAATATGGGATTTGGTGTGTGTGTGTGTGTGTGTGT 1243
QY 1017 ACCCAGAAATCTA---GACCTCTTCTATTATACAGCTGCAAAAATTTGAAATCTTTC 1074
DB 1244 ACCCAGAAATCTAGAGGATCTCTCTTTTATAACAATCAAGAAATATCAAGTACTTTC 1404
QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATTAATTTGTGTGTGTGTGT 1122
DB 1304 TCTTAATTTAATTTTCTGCTGTATGCAAGAAATAATTTGTGTGTGTGTGTGTGT 1454

RESULT 5
US-60-453-135-6918
: Sequence 6918, Application US/604533135
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: JAKOUHOVA, Olga
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
: CURRENT APPLICATION NUMBER: US/60/453,135
: CURRENT FILING DATE: 2003-03-10
: NUMBER OF SEQ ID NOS: 82762
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6918
: LENGTH: 1514
: TYPE: DNA
: ORGANISM: Homo sapiens
US-60-453-135-6918

Query Match      75.8%; Score 853.6; DB 12; Length 1514;
Best Local Similarity 86.6%; Pred No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATAGTATGGGAGCACCTGCGAGTTCTTTC---ATGAATGGCTAGTTGGGCTGC 57
DB 224 GGACTCTTACTCTGGGAGCACTGCTGTCTATCACAATGAAGCGCTGTTGTGCGC 283

QY 58 TTCCTGTGCTCTATGCGAGTGGCACAAAGTACATAAGATCCCACTCTGGATCATCACT 117
DB 284 TCTTGTGCTCTGCTGCGAGTGGCACATGCTGATTAAGATCCTACCTGGATCACT 343

QY 118 GGAATCTCTGGAGAAACCTACAGCAACAAT-CAAGGAAGAGAAATGAGGAAGTAGCAC 176
DB 344 GGCACTCTGGAGAAACCTATGGCAACAATCAAGGAAAGAAATGAAGAAGCAGTAC 403

QY 177 GFGCTCTATCTGGGAAACCTATTAATTTGTGATGCTTCACATCTGGACACTCAA 236
DB 404 GAGCTCTATCTGGGAAACCTATTAATTTGTGATGCTTCACACCTGGAGCACTCAA 463

QY 237 TGGGAATGCAATCATATGATCTAGGCATGAACCATCTGGGAGACATGCTGGTGAAGAG 296
DB 464 TGGGAATGCAATCATATGATCTAGGCATGAACCATCTGGGAGACATGCTGGTGAAGAG 523

QY 297 TGATATCTTTGATGGTTCCTTCAGAGTTCCTCAGAGCAATGGCAGAGAAATGTCATTATA 356
DB 524 TGATGCTTTGATGGTTCCTTCAGAGTTCCTCAGAGCAATGGCAGAGAAATGTCATTATA 583

QY 357 GGTCAACTCTAATCAGAAATTCCTGATCTGTGGACTGGAGAGAGAGGCTCTGTATA 416
DB 584 AGTCAAACTCTAATCAGAAATTCCTGATCTGTGGACTGGAGAGAGAGGCTCTGTATA 643

QY 417 CTGAAGTGAATACAGAGTTCCTTCAGAGTTCCTCAGAGCAATGGCAGAGAAATGTCATTATA 476
DB 644 CTGAAGTGAATACAGAGTTCCTTCAGAGTTCCTCAGAGCAATGGCAGAGAAATGTCATTATA 703

QY 477 TGGAGCAACTGGAAGCTAAAACAGGAAAGCTGGTGTCTGTAGTCAGACAGAACTGG 536
DB 704 TGGAGCAACTGGAAGCTAAAACAGGAAAGCTGGTGTCTGTAGTCAGACAGAACTGG 763

QY 537 TAGATGCTCAACTGAAAATATGGAATATAAGGCTGCAATGGCGGTTTCATGACAAGT 596
DB 764 TAGATGCTCAACTGAAAATATGGAATATAAGGCTGCAATGGCGGTTTCATGACAAGT 823

QY 597 CTTTCAGTATATTATGATACAGAGGCAATGATTCAGAGCTTCCTATCCCTACAAG 656
DB 824 CTTTCAGTATATTATGATACAGAGGCAATGATTCAGAGCTTCCTATCCCTACAAG 883

QY 657 CCATGAATGGAAGTGCAGATATGACTCAAAAAGGAGCTGCCACATGTTCAAAAGTATA 716
DB 884 CCATGAATGGAAGTGCAGATATGACTCAAAAAGGAGCTGCCACATGTTCAAAAGTATA 943

QY 717 CTGAATCTCCCTTTGGAGTGAAGATGCTTTAAAAGAAAGCTGTGTGCCCAATAAGAGGCTGC 776

```



```

QY 1017 ACCGAGAAATCTA---GACCTCTTCATTTTATAACAAGTCCAAAAATTTGAACACTTTC 1073
DB 1244 ACCGAGAAATCTAGAGATCTCTCTCTTTTATAACAATCAAGAAATATGAGCACTTTC 1303
QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA 1122
DB 1304 TCTTAATTTAATTTTCTGCTGTATCCAGAGAATAAATTTGTGTCATGA 1353

RESULT 7
US-60-455-444-3608
; Sequence 3608, Application US/60455444
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001455
; CURRENT APPLICATION NUMBER: US/60/455,444
; CURRENT FILING DATE: 2003-03-18
; NUMBER OF SEQ ID NOS: 50986
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3608
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-455-444-3608

Query Match 75.8%; Score 853.6; DB 13; Length 1514;
Best Local Similarity 86.6%; Pred. No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATTAGTATGGGAGCACCTGCAGGTTCTTTTC---ATGAAATGGCTAGTTGGGCTGC 57
DB 224 GCACCTCTTACTRGAGCAACTGCTGCTTCTATCACAATGAACGGCTGGTTGTGCGC 283

QY 58 TTCGTTGTGCTCTATGCAATGGCAGCAAGTACATCAATGATCTCTGATCATCAT 117
DB 284 TCTTGGTGTGCTCTCTGCTGCGGACAGATGTCATGAAGATCTTACCTGATCACCAC 343

QY 118 GGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAATGAGGAAGTAGC 176
DB 344 GCATCTCTGGAAGAAACCTATGCAACAATCAAGGAAGAAATGAGGAAGTAGC 403

QY 177 GCGCTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCACAACTGGAACACTCAA 236
DB 404 GACGCTCTCATCTCGGAAAAAATCTAAATTTGTGATGCTTCACAACTGGAACACTCAA 463

QY 237 TGGGAATGATTCATATGATCTAGGCAATGCAATCTGAGACATGCTGCTGAAGAAG 296
DB 464 TGGGAATGATTCATATGATCTAGGCAATGCAATCTGAGACATGCTGCTGAAGAAG 523

QY 297 TCATATCTTTGATGGTTCCCTGAGAGTTCCCGAGCAATGGCAGCAAAATGCTCACTTATA 356
DB 524 TGATGCTCTTGATGAGTTCCCTGAGAGTTCCCGAGCAATGGCAGCAAAATGCTCACTTATA 583

QY 357 GGTCAAACTCTAATCAGAAATTCGCTGATCTGCTGACTGGAGAGAGAGGCTGTGTTA 416
DB 584 AGTCAAACTCTAATCAGAAATTCGCTGATCTGCTGACTGGAGAGAGAGGCTGTGTTA 643

QY 417 CTGAAGTGAATATACCAAGGTTCTTGTGTGTTGTTGGCTTTTCAGCGCTTGGGGGCC 476
DB 644 CTGAAGTGAATATACCAAGGTTCTTGTGTGTTGTTGGCTTTTCAGCGCTTGGGGGCC 703

QY 477 TGAAGCACTGCTGAAGCTAAACAGGAAGCTGGTGTCTCTGATGCTGACAGAACTTGG 536
DB 704 TGAAGCACTGCTGAAGCTAAACAGGAAGCTGGTGTCTCTGATGCTGACAGAACTTGG 763

QY 537 TAGATTGCTCAACTCAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGACACTG 596
DB 764 TAGATTGCTCAACTCAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGACACTG 823

QY 597 CTTTCAGTATATTATTGATAACAAAGGCAATGATTTCAGAAAGCTTCTCTATCCCTACAAG 656

```

```

DB 824 CTTTCCAGTACATCATTGATAACAAGGCACTGACACTCAGACGCTTCTATCTCTACAAAG 884
QY 657 CCATGAATGGAAGTGCAGATATGACATCAAAAAGGCAAGCTGCAATGTTTAAATATA 716
DB 884 CCATGATAGGAATATGCAATATGATCAAAAATATGCTGCAATATGCAATATGCAAT 944
QY 717 CTGAACCTCCCTTTGGAGTGAAGATCAATTAANAAGGCTGCAATTAANAAGGCTG 776
DB 944 CTGAACCTCCCTTTGGAGTGAAGATCAATTAANAAGGCTGCAATTAANAAGGCTG 1004
QY 777 TGTCTGTGGTATAGATGAGGAGCAATATTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 846
DB 1004 TGTCTGTGGTATAGATGAGGAGCAATATTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 1064
QY 837 ATGAACATCTCTACTCAGAAATGCAATGATGATGATGATGATGATGATGATGATGAT 896
DB 1064 ATGAACATCTCTACTCAGAAATGCAATGATGATGATGATGATGATGATGATGATGAT 1124
QY 897 TTAATGGAAGAGTACTGCTGGAAGAGAGTGGGAGGCTGAGTGGGAGGCTGAGTGGG 956
DB 1124 TTAATGGAAGAGTACTGCTGGAAGAGAGTGGGAGGCTGAGTGGGAGGCTGAGTGGG 1184
QY 957 GATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1016
DB 1184 GATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1244
QY 1017 ACCGAGAAATCTA---GACCTCTTCATTTTATAACAAGTCCAAAAATTTGAACACTTTC 1074
DB 1244 ACCGAGAAATCTAGAGATCTCTCTCTTTTATAACAATCAAGAAATATGAGCACTTTC 1304
QY 1074 TCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA 1122
DB 1304 TCTTAATTTAATTTTCTGCTGTATCCAGAGAATAAATTTGTGTCATGA 1353

RESULT 8
US-60-465-241-3608
; Sequence 3608, Application US/60465241
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: BEGOVICH, Ann
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001468
; CURRENT APPLICATION NUMBER: US/60/465,241
; CURRENT FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 258418
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3608
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-465-241-3608

Query Match 75.8%; Score 853.6; DB 13; Length 1514;
Best Local Similarity 86.6%; Pred. No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATTATTAGTATGGGAGCACTGAGGTTCTTTTC---ATGAAATGGCTAGTTGGGCTGC 57
DB 224 GCATCTTACTGTTGAGAGCAACTGCTGCTTCTATCACAATGAACGGCTGGTTGTGCGC 284
QY 58 TTCGTTGTGCTCTATGCAATGGCAGCAAGTACATCAATGATCTCTGATCATCAT 117
DB 284 TCTTGGTGTGCTCTCTGCTGCGGACAGATGTCATGAAGATCTTACCTGATCACCAC 344
QY 118 GGAATCTCTGGAAGAAACCTACAGCAACAAT-CAAGGAAGAGAAATGAGGAAGTAGC 176
DB 344 GCATCTCTGGAAGAAACCTATGCAACAATCAAGGAAGAAATGAGGAAGTAGC 404
QY 177 GCGCTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCACAACTGGAACACTCAA 236
DB 404 GACGCTCTCATCTCGGAAAAAATCTAAATTTGTGATGCTTCACAACTGGAACACTCAA 463
QY 237 TGGGAATGATTCATATGATCTAGGCAATGCAATCTGAGACATGCTGCTGAAGAAG 296
DB 464 TGGGAATGATTCATATGATCTAGGCAATGCAATCTGAGACATGCTGCTGAAGAAG 523
QY 297 TCATATCTTTGATGGTTCCCTGAGAGTTCCCGAGCAATGGCAGCAAAATGCTCACTTATA 356
DB 524 TGATGCTCTTGATGAGTTCCCTGAGAGTTCCCGAGCAATGGCAGCAAAATGCTCACTTATA 583
QY 357 GGTCAAACTCTAATCAGAAATTCGCTGATCTGCTGACTGGAGAGAGAGGCTGTGTTA 416
DB 584 AGTCAAACTCTAATCAGAAATTCGCTGATCTGCTGACTGGAGAGAGAGGCTGTGTTA 643
QY 417 CTGAAGTGAATATACCAAGGTTCTTGTGTGTTGTTGGCTTTTCAGCGCTTGGGGGCC 476
DB 644 CTGAAGTGAATATACCAAGGTTCTTGTGTGTTGTTGGCTTTTCAGCGCTTGGGGGCC 703
QY 477 TGAAGCACTGCTGAAGCTAAACAGGAAGCTGGTGTCTCTGATGCTGACAGAACTTGG 536
DB 704 TGAAGCACTGCTGAAGCTAAACAGGAAGCTGGTGTCTCTGATGCTGACAGAACTTGG 763
QY 537 TAGATTGCTCAACTCAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGACACTG 596
DB 764 TAGATTGCTCAACTCAAAATATGGAATTAAGGCTGCAATGGCGGTTTTCATGACACTG 823
QY 597 CTTTCAGTATATTATTGATAACAAAGGCAATGATTTCAGAAAGCTTCTCTATCCCTACAAG 656

```



```

404 CACGTCATCTCGGAAAGAAATCTAAAGTTTGTGATGCTTCAAACTGGAGCAATCAA 463
QY
237 TGGGAATGCTATGATCTAGGCATGAACATCTGGGAGACATGACTGGTGAAGAAG 296
Db
464 TGGGAATGCTATGATCTAGGCATGAACATCTGGGAGACATGACTGGTGAAGAAG 523
QY
297 TGATATCTTTTCATGGCTTCCCTGAGAGTTCCCAAGCAATGGCAGAGAAATGTCATATA 356
Db
524 TGATGCTTTGAGAGTTCCCTGAGAGTTCCCAAGCAATGGCAGAGAAATGTCATATA 583
QY
357 GGTCAACTCTAATCAGAAATGCTGATCTGGGAGACATGACTGGTGAAGAAG 416
Db
584 AGTCAAACTCTAATCAGAAATGCTGATCTGGGAGACATGACTGGTGAAGAAG 643
QY
417 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 476
Db
644 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 703
QY
477 TGGAGCAACTGTAAGCTAATAACAGCAAGAGCTGCTCTCTGAGTGCACAGAACTGG 536
Db
704 TGGAGCAACTGTAAGCTAATAACAGCAAGAGCTGCTCTCTGAGTGCACAGAACTGG 763
QY
537 TAGATGCTCAACTGAAATATGGAATATGGAATATGGAATATGGAATATGGAATATG 596
Db
764 TGGATGCTCAACTGAAATATGGAATATGGAATATGGAATATGGAATATGGAATATG 823
QY
597 TTTTCAGTATATATGATACACAGGCTATGATTCAGAGCTTCTCTGAGTGCACAGAA 656
Db
824 CTTCAGTATATATGATACACAGGCTATGATTCAGAGCTTCTCTGAGTGCACAGAA 883
QY
657 CCAATGAATGGAAGTGCAGATATGATTCAGAGCTTCTCTGAGTGCACAGAACTGG 716
Db
884 CCAATGAATGGAAGTGCAGATATGATTCAGAGCTTCTCTGAGTGCACAGAACTGG 943
QY
717 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 776
Db
944 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 1003
QY
777 TGTCTGCTATAGTGGAGCTATGCTTCTTCTCTGAGTGCACAGAACTGGTCTACT 836
Db
1004 TGTCTGCTATAGTGGAGCTATGCTTCTTCTCTGAGTGCACAGAACTGGTCTACT 1063
QY
837 ATGAACATCTCTACTCAGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 896
Db
1064 ATGAACATCTCTACTCAGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1123
QY
897 TTAATGGGAAGACTACTGCTTGTGAAAGAGCTGGGGCTCAACTTTGGTGACCAAG 956
Db
1124 TTAATGGGAAGACTACTGCTTGTGAAAGAGCTGGGGCTCAACTTTGGTGACCAAG 1183
QY
957 GATATATTCGATGGGAGAAACAGTGGAAATATGCTGGGATGCTAGTTATGCTCTTT 1016
Db
1184 GATATATTCGATGGGAGAAACAGTGGAAATATGCTGGGATGCTAGTTATGCTCTTT 1243
QY
1017 ACCAGAAATCTA--GACCTCTCATTTTATACAAAGTCCAAATAATGGAACACTTTC 1073
Db
1244 ACCAGAAATCTAGAGATCTCTCTTTTATACAAATGCAATATGGAACACTTTC 1303
QY
1074 TCTTAATTAATTTTACTGCTATA-ATAGTAAATAATGATGTCATGA 1122
Db
1304 TCTTAATTAATTTTCTCTGCTATCCAGAAATATGCTGTCATGA 1353

```

```

RESULT 9
US-60-466-412-6918
; Sequence 6918, Application US/60466412
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michelle
; APPLICANT: IAKUBOVA, Olga
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001466
; CURRENT APPLICATION NUMBER: US/60466.412

```

```

; CURRENT FILING DATE: 2003-04-30
; NUMBER OF SEQ IN NOS: 429241
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6918
; LENGTH: 1514
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-466-412-6918

```

```

Query Match 75.8%; Score 853.6; DB 13; Length 1514;
Best Local Similarity 86.6%; Pred. No. 5.6e-220;
Matches 979; Conservative 7; Mismatches 136; Indels 8; Gaps 4;

QY 1 GCATATTAGTATGGGAGCACTCGAGTTCCTTTC---ATGAAATGGCTAGTTGGCGTGC 57
Db 224 GGACTCTTACTGTRGGAGCAACTGCTGTTCTATCAACATGMAAGCGCTGGTTGTGCGC 283
QY 58 TTCCGTTGCTCCTATGCACTGGCACAAGTACATATAAGATCCACACTCTCGATCATCACT 117
Db 284 TCTTGTGCTGCTCTCTGCTGTCAGTGGCAGATTCGATAAAGATCCTACCTGGATCCACACT 343
QY 118 GGAATCTCTGGAAGAAACCTACAGCAAAACAT-CAAGGAAGAGAAATGAGAAAGTATGAC 176
Db 344 GGATCTCTGGAAGAAACCTATGCAAAACATACAGGAAGAAATGAGAAAGTATGAC 403
QY 177 GGCCTCTCATCTGGGAAAAAATCTAAATTTGATGCTTCACAATCTGGAACACTCAA 236
Db 404 GAGCTCTCATCTGGGAAAAAATCTAAATTTGATGCTTCACAATCTGGAACACTCAA 463
QY 237 TGGGAATGCAATTCATATGATCTAGGCATCAACCACTCGGAGAGACATCTCTCAACAAG 296
Db 464 TGGGAATGCAATTCATATGATCTAGGCATCAACCACTCGGAGAGACATGACCACTCAACAAG 523
QY 297 TGATATCTTTGATGGGTTCCCTGAGAGTTCCCAAGCAATGCGAGAGAAATGCTCTATA 356
Db 524 TGATGCTTTGAGAGTTCCCTGAGAGTTCCCAAGCAATGCGAGAGAAATATCACATATA 583
QY 357 GGTCAAACTCTAATCAGAAATTCCTGATTCCTGAGTGCAGAGAGAGGCTCTGTATA 416
Db 584 AGTCAAACTCTAATGAGATTTGCTGATTCCTGAGTGCAGAGAGAGGCTCTGTATA 643
QY 417 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 476
Db 644 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 703
QY 477 TGGAGCAACTGTAAGCTAATAACAGCAAGAGCTGCTCTCTGAGTGCACAGAACTGG 536
Db 704 TGGAGCAACTGTAAGCTAATAACAGCAAGAGCTGCTCTCTGAGTGCACAGAACTGG 763
QY 537 TAGATGCTCAACTGAAATATGGAATATGGAATATGGAATATGGAATATGGAATATG 596
Db 764 TGGATGCTCAACTGAAATATGGAATATGGAATATGGAATATGGAATATGGAATATG 823
QY 597 TTTTCAGTATATATGATACACAGGCTATGATTCAGAGCTTCTCTGAGTGCACAGAA 656
Db 824 CTTCAGTATATATGATACACAGGCTATGATTCAGAGCTTCTCTGAGTGCACAGAA 883
QY 657 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 716
Db 884 CCAATGAATGGAAGTGCAGATATGATTCAGAGCTTCTCTGAGTGCACAGAACTGG 943
QY 717 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 776
Db 944 CTGAAGTGAATACCAAGGTTCTTGTGTGCTTGTGGGCTTTTCAGGCGCTGGGGGCC 1003
QY 777 TGTCTGCTATAGTGGAGCTATGCTTCTTCTCTGAGTGCACAGAACTGGTCTACT 836
Db 1004 TGTCTGCTATAGTGGAGCTATGCTTCTTCTCTGAGTGCACAGAACTGGTCTACT 1063
QY 837 ATGAACATCTCTACTCAGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 896
Db 1064 ATGAACATCTCTACTCAGAAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1123

```

Qy	897	TTAATGGGAAGACTACTTGGCTTGTGTAAAAACAGCTGGGGCCCTCAACTTTGGTGTGACCAAG	956
Db	1124	TTAATGGGAAGAAATACTTGGCTTGTAAAAAACAGCTGGGGCCCAACACTTTGGTGTGAAGAAG	1183
Qy	957	GATATATTCGGATTGGCAAGAAACAGTGGAAATCACCTCTGGGATTGCTAGTTATCCCTCTT	1016
Db	1184	GATATATTCGGATTGGCAAGAAATTAAGGAATCATTTGTGGGATTGTAGCTTTCCCTCTT	1243
Qy	1017	ACCAGAAATCTA---GACCTCTTCATTTTATAACAAAGTCCAAAAATTAAGAAACACTTTC	1073
Db	1244	ACCAGAAATCTAGAGGATCTCTCCITTTTATAACAAATCAAGAAATATGAAGCACTTTC	1303
Qy	1074	TCCTTAATTTAATTTTACCTGCTATA-ATAGTAAAAATAAATGTGTCATGA	1122
Db	1304	TCCTTAACCTTAATTTTTCCTGCTATCCAGAGAAATAATTTGTGTCATGA	1353

RESULT 10

```

US-09-873-367C-173
: Sequence 173, Application US/09873367C
: GENERAL INFORMATION:
: APPLICANT: Young, Paul
: APPLICANT: Soppet, Daniel
: APPLICANT: Endress, Gregory
: APPLICANT: Augustus, Meena
: APPLICANT: Ebner, Reinhard
: APPLICANT: Carter, Kenneth
: TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
: TITLE OF INVENTION: Signature Gene Sets
: FILE REFERENCE: 689290_64
: CURRENT APPLICATION NUMBER: US/09/873,367C
: CURRENT FILING DATE: 2003-04-29
: PRIOR APPLICATION NUMBER: U.S. 60/236,891
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: U.S. 60/236,842
: PRIOR FILING DATE: 2000-09-29
: PRIOR APPLICATION NUMBER: U.S. 60/244,867
: PRIOR FILING DATE: 2000-11-01
: PRIOR APPLICATION NUMBER: U.S. 60/245,084
: PRIOR FILING DATE: 2000-11-01
: NUMBER OF SEQ ID NOS: 1067
: SOFTWARE: Patent In version 3.0
: SEQ ID NO 173
: LENGTH: 1763
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-873-367C-173

```

Query Match	75.18	Score 845.4	DB 5	Length 1763
Best Local Similarity	87.28	Pred. No. 9.5e-218,		
Matches 986: Conservative	0	Mismatches 136:	Indels 9:	Gaps 5

QY	1	GCATTATTAGTATGGGACACCTCGAGTTCTTTTC---ATGAAATGGCTAGTTGGCTGC	57
Db	99	GGACTCTTACTGTAGGACAACCTGCTGGTCTATCAATGAAACGGCTGGTTTGTGTGC	158
QY	58	TTCCGPTGTGTCCTATGCACTGGCACAAGTACATAAAGATCCACTCGGATCATCACT	117
Db	159	TCTTGGTGTGTCCTCGCACTGGCACAGTTGGATAAAGATCCTACCTTGGATCACCCT	218
QY	118	GGAAATCTCTGGAAAGAAACCTACAGAACAAAT-CAAGGAAGAGATAGGAGATAGCAC	176
Db	219	GGCATCTCTGGGAAGAAACCTATGGCAACAATACAGGAAAGAAATGAAGAGCAGTAC	278
QY	177	GGCGTCTCATCTGGCAAAAAAATCTAAAAATTGTGATGCTTCACAATCTGGGAACACTCAA	236
Db	279	GACGTCTCATCTGGGAAAGAAATCTAAAGTTTGTGTGCTTCACAACTGGGACATTCAA	338
QY	237	TGGGAATGCATTCATATCATCTAGGCATGAACCATCTGGGAGACATGACTGGTGAAGAG	296
Db	339	TGGGAATGCATTCATACCATCTGGGCATGAACCACTGGGAGACATGACCAGCTGAAGAG	398
QY	297	TGATATCTTTATGAGTTTCTTGASAGTTCCAGAGCAATGCGAGAGAAATGTCACTTATA	356

Db 219 GGCACTCTGGAGAAAGCACTATGGTAAACAAATAAAGGAAAGAAATGAAGAAGCAGTAC 278
QY 177 GGCTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCACAACTCTGGAACACTCAA 236
Db 279 GACGCTCATCTGGGAAAAAATCTAAAGTTTGTGATGCTTCACAACTCTGGAGCAATCAA 338
QY 237 TGGGAATGATTATATGATCTAGGCATGAACCATCTGGGAGACATGACTGGTGAAGAAG 296
Db 339 TGGGAATGCATCATACGATCTGGGCATCAACCACTGGGACACATGACCACTGAAGAAG 398
QY 297 TGATATCTTTGATGGTTCCCTGAGAGTTCCAGGCCAATGGCAGAGAAATGTCACCTATA 356
Db 399 TGATGCTTTGATGATTTCCCTGAGAGTTCCAGGCCAATGGCAGAGAAATATCACATATA 458
QY 357 GGTCAAACTCTAATCAGAAATTCCTGATTCTGTGACTGGAGAGAGAGGGCTGTGTTA 416
Db 459 AGTCAAACTCTAATCGGATATTGCCGTGATTTCTGGACTGGAGAGAGAGGGTGTGTTA 518
QY 417 CTGAAGTGAATATACAGGTTCTTGTGTTGCTTTGTGGCTTTTCAGCGCTTTGGGGGCC 476
Db 519 CTGAAGTGAATATCAAGTCTTGTGTTGCTTTGTGGCTTTTCAGTGTCTGGGGGCC 578
QY 477 TCGAAGCACAACCTGAAGCTTAAAAACAGGAAAGCTGGTGTCTCTGACTGCACAGAACTTGG 536
Db 579 TGGAGCAGAGCTGAAGCTGAAGCTGAAGCAGGAAAGCTGGTGTCTCTGAGTGCCCAACCTGG 638
QY 537 TAGATTGCTCAACTGAAAAATATGGGAATAAAGGCTGCAATGGCGGTTTCATCACAACCTG 596
Db 639 TGGATTGCTCAACTGAAAAATATGGAAACAAGGCTGCAATGGTGGCTTCATCACAACGG 698
QY 597 CTTTCAGTATATTATTGATACAGGGCATTTGATTGAGAGCTTCCTATCCCTACAAAG 656
Db 699 CTTTCAGTATACATATGATAACAGGGCATCGACTCAGACGCTTCCTATCCCTACAAAG 758
QY 657 CCATGAATGGAAGTCACATATGACTCAAAAAGCGAGCTGCCACATCTTCCAAAGTATA 716
Db 759 CCATGAATCAGAAATGTCAATATGACTCAAAATATCGTGTGCCACATGTTCAAAAGTACA 818
QY 717 CTGAACCTCCCTTTGGCAGTGAAGATGCCCTTAAAGAGAGCTGGCCCAATAAAGGACCTG 776
Db 819 CTGAACCTCCCTTATGCCAGAGAAGATGCTCTGAAAGAGCTGTGCCCAATAAAGGCCAG 878
QY 777 TGTCTGTGCTATAGATGGAGCCACTATTCTTCTCTGTACAGAGTGGTGTCTACT 836
Db 879 TGTCTGTGGTGTAGATGGCGCTCATCCCTCTTCTTCTCTACAGAGTGGTGTCTACT 938
QY 837 ATGAACCATCCTCTACTCAGAAATGTAATCATCGAGTATTAGTGTGGCTATGCTTAACC 896
Db 939 ATGAACCATCCTGTACTCAGAAATGTAATCATCGTGTACTTGTGTGGTATGCTGATC 998
QY 897 TTAATGGGAAAGACTACTGGCTTGTGAAAGAGAGCTGGGGCTCAACTTTGGTGACCAAG 956
Db 999 TTAATGGGAAAGAAATACTGGCTTGTGAAAGAGAGCTGGGGCCACAACTTTGGTGAAGAAG 1058
QY 957 GATATATTCGGATGGCAAGAAACAGTGGAAATCACTGTGGGATTTGCTAGTTATCCCTCTT 1016
Db 1059 GATATATTCGGATGGCAAGAAATCAAGGAAATCATTGTGGGATTTGCTAGCTTCCCTCTT 1118
QY 1017 ACCCAGAAATCTA---GACCTCTCATTTTATAACAAGTCCAAAAAATTTGAACAACTTTC 1073
Db 1119 ACCCAGAAATCTAGAGGATCTCTCCTTTTATAACAATAATCAAGAAATATAGAAAAATATA 1178
QY 1074 TCTTAATT 1081
Db 1179 GTTTGATT 1186

Search completed: July 1, 2003, 12:00:03
Job time : 989 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 11:05:43 ; Search time 72 Seconds

(without alignments)
4167.783 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1997

Sequence: 1 gcattattagtagggagca.....aataaatgttcacgacct 1126

Scoring table:

BLOSUM62
Xgapop 10 0, Xgapext 0.5
Ygapop 10 0, Ygapext 0.5
Fgapop 6 0, Fgapext 7.0
Delop 6 0, Delext 7.0

Searched: 908470 seqs, 13250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DH seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+_{n2p}.model -DEV-xlp
-O=/cpn2_1/USPTU_spool/US10010580/runat_27062003_104535_11686/app_query.fasta.1.1287
-DH=A_Genseq_101002 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT=plp -NORW-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10010580 -RUN_1_114 -runat_27062003_104535_11686 -NCPU=6 -ICPU=3
-NO_MMWP -LARGEQUERY -NEW_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCAPUP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

1: /SID52/gcadata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcadata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcadata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcadata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcadata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcadata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcadata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcadata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcadata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcadata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcadata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcadata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcadata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcadata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcadata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcadata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcadata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcadata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcadata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcadata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcadata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcadata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcadata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	1569	78.6	331	21	AAV59634	Human Cathepsin S
2	940.5	47.1	329	21	AAH30506	A monkey cathepsin
3	939.5	47.0	329	16	AAH82720	Human osteoclast-d
4	939.5	47.0	329	17	AAH95599	Prepro-cathepsin-O
5	939.5	47.0	329	19	AAH94216	Human cathepsin K
6	939.5	47.0	329	19	AAH41645	Human cathepsin K
7	939.5	47.0	329	21	AAH81000	Human procathepsin
8	939.5	47.0	329	21	AAH57115	Human prostate can
9	937.5	46.9	329	19	AAH70701	Rat Cathepsin K po
10	932.5	46.7	329	19	AAH41984	Human cathepsin X
11	932.5	46.7	329	23	AAH81810	Human cathepsin X
12	930.5	46.6	329	19	AAH48947	Mutant human cathe
13	930.5	46.6	329	19	AAH48939	Mutant human cathe
14	929.5	46.5	329	19	AAH48946	Mutant human cathe
15	929.5	46.5	329	19	AAH48938	Mutant human cathe
16	925.5	46.3	329	19	AAH42102	Mus musculus cathe
17	922.5	46.2	329	23	AAH57267	Mouse ischaemic co
18	915	45.8	329	19	AAH44778	Human recombinant
19	901	45.1	322	22	AAH66260	Human recombinant
20	843	42.2	334	20	AAH02358	Polypeptide identi
21	843	42.2	334	22	AAH81487	Human cathepsin I2
22	843	42.2	334	22	AAU12177	Human pro305 poly
23	832	41.7	334	22	AAH94300	Human cathepsin (L
24	812.5	40.7	336	22	ABG21426	Novel human diagno
25	808	40.5	333	22	AAH93531	Human polypeptide,
26	806	40.4	333	19	AAH47031	Human procathepsin
27	806	40.4	333	23	ABH77396	Human cathepsin I.
28	805.5	40.3	333	23	ABG66692	Human novel polype
29	798.5	40.0	333	23	ABG66714	Human novel polype
30	793	39.7	333	9	AAH28829	Human procathepsin
31	792	39.7	333	9	AAH2094	phu-16, sequence en
32	767	38.4	341	22	ABH62597	Drosophila melanog
33	738.5	37.0	333	23	ABH98883	Human protease PKT
34	722	36.2	333	23	ABG61771	Novel cathepsin-L
35	709.5	35.5	326	15	AAH57071	Fasciola hepatica
36	705	35.3	419	21	AAH35414	zebra mays protein f
37	705	35.3	424	21	AAH35413	zebra mays protein f
38	701	35.1	330	23	AAH09882	Amino acid sequenc
39	689.5	34.5	461	20	AAH89558	Triticum sp. cyste
40	686.5	34.4	326	15	AAH57080	Fasciola hepatica
41	681	34.1	330	21	AAH13420	Marine sponge sili
42	665	33.3	211	19	AAH44779	Human cathepsin K
43	651	32.6	472	20	AAH89559	Triticum sp. cyste
44	649.5	32.5	500	24	AAH65766	Cysteine protease
45	649	32.5	463	18	AAH19541	Soybean thiol prot

ALIGNMENTS

RESULT 1
AAV59634
ID AAV59634 standard; Protein; 331 AA.
XX
AC AAV59634;
XX
DT 27-MAR-2000 (first entry)
XX
DE Human Cathepsin S amino acid sequence.
DE
DE Cathepsin S; human; endometriosi; treatment; diagnose.
KW
XX Homo sapiens.
XX WO963115-A2.
XX
XX
PD 09-DEC-1999.
XX
PF 03-JUN-1999; 99WO-US12335.

XX 04 -JUN-1998; 98US-0086017.
 XX (REPR-) REPROGEN INC.
 XX PI Schneider P, Yamamoto KK, French CK;
 XX DR WPI: 2000-086986/07.
 XX DR N-PSDR; AA756150
 XX
 PT Use of cathepsin S in the diagnosis and treatment of endometriosis -
 XX Examples: Page 15; 60pp; English.
 XX
 CC This is the human cathepsin S protein sequence. Detecting levels of the
 CC cathepsin S gene product in a sample compared to a control sample can be
 CC used as a method for diagnosing endometriosis. The invention also relates
 CC to a method for treating endometriosis through the administration of a
 CC probe comprising a detectable label and a ligand that specifically binds
 CC a cathepsin S gene product to the subject. The endometriotic lesion can
 CC be identified in situ by locating bound labelled probe; and the lesion
 CC can be excised. The methods are useful to diagnose, monitor the progress
 CC of and treat endometriosis in a subject. The methods are also useful for
 CC screening for modulators of cathepsin S gene production in endometrial
 CC cells. Antisense cathepsin S gene oligonucleotides are useful for the
 CC treatment of endometriosis by down-regulating cathepsin S genes.
 XX
 XX Sequence 331 AA:
 SQ
 Alignment Scores:
 Pred. No.: 3,77e-165 Length: 331
 Score: 1569.00 Matches: 290
 Percent Similarity: 93.35% Conservative: 19
 Best local Similarity: 87.61% Mismatches: 22
 Query Match: 78.57% Indels: 1
 DB: 21 Gaps: 0

US-10-010-580-1 (1-1126) x ANYS9634 (1-331)

QY 36 ATGAATGGCTAGTTGGCTCTCTCCGTTGCTCTCTATGCGTGGCAGCAAGTACATAAA 95
 DB 1 MetLysArgLeuValCysValLeuLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
 QY 96 GATCCACCTCTGATCATCATCTGGAATCTCTGGAAGAAACCTACACCAACATC-AAG 154
 DB 21 AspProThrLeuAspHisHisThrHisLeuTrpLysThrGlyLysGlnTrpLys 40
 QY 155 GAAAGAGATGAGGAAGTAGGAAGGCTCTCTATCTGGGAAAAAATCTAAATTTTGATG 214
 DB 41 GluLysAsnGluGluAlaValArgArgLeuLeuTrpGluLysAsnLeuLysPheValMet 60
 QY 215 CTTCACATCTGGACACTCAATGGGAATGCAATTCATATCATATCAATGAGCATGACCATCTG 274
 DB 61 LeuHisAsnLeuGluHisSerMetGlyMetHisSerTyrAspLeuGlyMetAsnHisLeu 80
 QY 275 GSAACATCAATGATGAAAGTATATCTTTTAAAGAGTTCAGTAGTTCCTCCAGCCAA 334
 DB 81 GlyAspMetThrSerGluGluValMetSerLeuMetSerSerLeuArgValProSerGln 100
 QY 335 TGAAGAAGAAATATATTTATATAGTAAATTTAAATAGAAATTTGATTTGTGGAC 394
 DB 101 TrpGlnArgAsnLeuThrTyrLysSerAsnProAsnArgLeuLeuProAspSerValAsp 120
 QY 395 TGGACAGACAGCGCTGTTACTCAACTGAAATACCGCGTCTCTGTTGGTCTTGTGG 454
 DB 121 TrpArgGluLysGlyCysValThrGluValLysTyrGlnGlySerCysGlyAlaCysTrp 140
 QY 455 GCTTTCAGCGCTGAGGAGGAGTTCGAGAGCAATTAAGTAAAGTAAAGTAAAGTAAAGT 514
 DB 141 AlapheSerAlaValGlyAlaLeuGluAlaGlnLeuLysLeuLysThrGlyLysLeuVal 160
 QY 515 TCTCTGAGTGAAGAAATTTGATAGTGGTAAATTTAAATATATGGAATAAGGCTGC 574

DB 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGluLysTyrGlyAsnLysGlyCys 180
 QY 575 AATGGCGGCTTTTAAGACAACTGGCTTTCACAGATAATATATCAATAAATGATATATCA 634
 DB 181 AsnGlyClyPheMetThrThrAlapheGlnTyrIleIleAspAsnLysGlyIleAspSer 200
 QY 635 GAAATTTTAT 694
 DB 201 ASPALASerTyrProTyrLysAlaMetAspLysCysCysIleTyrAspSerLysTyrArg 220
 QY 695 GCTGGCACATGTTTAAAGATATATGAACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 754
 DB 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu 240
 QY 755 GCTGTGGCAATAAAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 814
 DB 241 AlaValAlaAsnLysGlyProValSerValGlyValAlaAspAlaArgHisProSerPhePhe 260
 QY 815 CTGTACAGAAAGTGTCTTACTATTAATATATATATATATATATATATATATATATATAT 874
 DB 261 LeuTyrArgSerGlyValTyrTyrGluProSerCysThrGlnAsnValAsnHisGlyVal 280
 QY 875 TTACTGCTGGCTATGCTAACCTTAATGGCAAGACTACTGCTTCTGCTGCTGCTGCTGCTG 934
 DB 281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGlyTyrTrpLeuValLysAsnSerTrp 400
 QY 935 GGCCTCAACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 994
 DB 301 GlyHisAsnPheGlyGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisLys 420
 QY 445 GAAATTTTAT 1027
 DB 321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 2

AAB30506 standard; Protein: 329 AA.

XX AAB30506;

DT 06-MAK-2001 (first entry)

DE A monkey cathepsin K polypeptide.

XX Monkey; cathepsin K; cysteine protease, osteoclast, bone resorption;
 KW osteoporosis; periodontal disease; arthritis; cancer; tumour metastasis;
 KW Paget's disease; bone disease.

OS Macaca fascicularis.

PH Key Location/Qualifiers

FT Misc-difference 143 /note- "Asn encoded by AGC"

FT Misc-difference 175 /note- "Ala encoded by GAT"

XX W0200061785-A2.

XX 19-OCT-2000.

XX 14-APR-2000; 2000W0-0510160.

XX 14-APR-1999; 99US-0129206.

XX 13-APR-2000; 2000US-0548512.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Field J;

XX WPI: 2000-665138/64.

XX N-PSDR; AAC62064.

PT Novel monkey cathepsin K polypeptide for screening for agonist and

US-10-010-580-1 (1-1126) x AAR95599 (1-329)

QY	944	TTTGSTGACCAAGSATATATTGSGATGKCAAGAAACAGTGGAAATCACTGIGGGATTGCT	100
Db	302	TRPGlyAsnLysGlyTrilleLeuMeCAlaIaGAsnLysAsnAlaCysGlyIleAla	321
QY	1004	AGTATATCCCTCTTACCACAGAAATC 1027	
Db	322	AsnLeuAlaSerPheProLysMeI 329	
RESULT 5			
AAW39216		AAW39216 standard; Protein; 329 AA.	
XX	AC	AAW39216;	
XX	DT	20-JUL-1998 (first entry)	
XX	DE	Human cathepsin K.	
XX	KW	Cathepsin K; human; osteoporosis; periodontal disease;	
KW	KW	Paget's disease; Gaucher's disease; Alzheimer's disease;	
KW	KW	central nervous system inflammation; hyperparathyroidism;	
KW	KW	bone degradation; dental implant degradation; metastasis; tumour;	
XX	XX	diagnosis; therapy.	
OS	XX	Homo sapiens.	
XX	PN	EP812916-A2.	
XX	PD	17-DEC-1997.	
XX	PF	19-MAY-1997; 97EP-0303395.	
XX	PR	26-AUG-1996; 96US-0026083.	
PR	PR	14-JUN-1996; 96US-0019942.	
PR	PR	17-JUN-1996; 96US-0020273.	
PR	PR	26-AUG-1996; 96WO-US14026.	
XX	XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	PA	(GENO-) INST GENOMIC RES.	
PA	PA	(SMIK) SMITHKLINE BEECHAM CORP.	
XX	PI	Adams MD, Blake JA, Debouck CM, Drake FH, Fitzgerald LM;	
PI	PI	Fraser CM, Gowen M, Hastings GA, Kirkness EF, Lee NH;	
PI	PI	Rood J;	
XX	DR	WPI: 1998-034977/04.	
DR	DR	N-PSDB; AAV09660.	
XX	PT	DNA encoding human cathepsin K - useful for diagnosing and treating	
PT	PT	diseases associated with cathepsin K e.g. osteoporosis, bone	
PT	PT	degradation, metastatic tumours, etc	
XX	PS	Claim 2; Page 46-47; 84pp; English.	
PS	XX	This polypeptide comprises human cathepsin K. the amino acid	
CC	CC	sequence can be deduced from isolated cathepsin K genomic DNA	
CC	CC	(see AAV09660). Cathepsin K polypeptides, including polypeptides	
CC	CC	or by differentially spliced polynucleotides of the gene, can be	
CC	CC	expressed in host cells and used to identify compounds which bind	
CC	CC	to and inhibit activation of cathepsin K. Antagonists may be	
CC	CC	useful in the treatment of osteoporosis, periodontal disease,	
CC	CC	Paget's disease, Gaucher's disease, central nervous system (CNS)	
CC	CC	inflammation, Alzheimer's disease, hyperparathyroidism, bone	
CC	CC	degradation, metastatic tumours, and degradation of bone implants	
CC	CC	and prostheses, especially dental implants. Assays for the	
CC	CC	presence of cathepsin K polypeptides can be used for the diagnosis	
CC	CC	of diseases characterised by aberrant expression of a cathepsin K	
CC	CC	polypeptide. The polypeptides can also be used to treat patients	
XX	XX	having need of cathepsin K.	
XX	XX	Sequence 329 AA;	

Score: 939.50 Matches: 183
 Percent Similarity: 70.438 Conservative: 48
 Best Local Similarity: 55.798 Mismatches: 88
 Query Match: 47.05% Indels: 10
 DB: 21 Gaps: 4

US-10-010-580-1 (1-1126) x AAM7071 (1-329)

```

QY 54 CTGCTCCCTGTCCTCTATGTCAGTGGCACAAGTACATTAAGATCCACTGTGATCAT 113
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 40 LeuLeuProValSerPheLea-----LeuTyrProGluGluLeuLeuSptThr 56
QY 114 CACTGCAATCTGCGAAGAAACCTACAGCAAAACATCAAGAGACAGAT--GAGGAAGTA 172
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 57 HISTRGLeuLeuTyrPheLysThrHisAlaGlySerGlnTyrAsnAsnLysValAspLeuIle 76
QY 173 GCAAGGCGTCTATCTGGAAAAAAATTTAAATTTGTGATGCTTCAATCTGGAAACAC 232
    .....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 77 SerArgArgLeuLeuIleProGluLysAsnLeuLysTyrIleSerIleHisAsnLeuGluAla 96
QY 233 TTAATGGAAATTCATTCATATATATATAGGATGATGATGATGATGATGATGATGAT 292
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 97 SerLeuGlyValHisThrTyrGluLeuAlaMetAsnHisLeuGlyAspMetThrSerGlu 116
QY 293 GAAGTATATCTTGTGATGGTCCCTGAGAGTTCCAGCCCAATGGCAGACA-----AAT 346
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 117 GluValValGlnLysMetThrGlyLeuLysValProLeuSerHisSerArgSerAsnArg 136
QY 347 GTGACTTATATAGTCAAACTATACAGAAATGGCTGATTTGTGTGATGGAGAGAGAG 406
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 137 ThrLeuTyrIleProGluTyrProGluLysArgAlaProAspSerValAspTyrArgLysLys 156
QY 407 GGCTGTGTACTGAAATGCAAAATTCAGAGGTCTTGTGTGCTGTTGGCTTTTCAAGCTGT 466
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 157 GlyTyrValThrProValLysAsnGlnGlyGlnGlySerGlySerArgLysSerSer 176
QY 467 GTGGGGCCCTGGAACACACTGAAAGCTAAAAACAGAAAGCTGGTCTCTGAGTGA 526
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 177 ValGlyValAlaLeuGlnLysLeuLysLysLysLysLysLysLysLysLysLysLysLys 196
QY 527 CCAACATCTGATGATCTCTCACTCAATGAAATATGAAATGAAATGAAATGAAATGAAATG 586
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 197 GluAsnLeuValAspCysValSerGlu-----AsnAspLysGlyGlyGlyGlyTyr 213
QY 587 ATGACAACCTGCTTCCAGTATATATATATATATATATATATATATATATATATATAT 646
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 214 MetThrAsnAlaPheGlnTyrValGlnLysAsnArgGlyIleAspSerGluAspAlaTyr 233
QY 647 CCTTACAACCCATGATGGAAGCTGACATATACCTCAAAAACGAGAGCTGCCACATGT 706
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 234 ProTyrValGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 253
QY 707 TCAAGATATATATATATATATATATATATATATATATATATATATATATATATATAT 766
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 254 ArgGlyTyrArgGlyLysLeuProGluLysGlnLysGlnLysGlnLysGlnLysGlnLys 273
QY 767 AAAGGAATATATATATATATATATATATATATATATATATATATATATATATATATAT 826
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 274 ValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheGlnPheTyrSerLys 293
QY 827 GGTGTCTACATGACCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 883
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 294 GluValAlaTyrTyrArgLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 313
QY 884 GATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 943
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 314 GlnTyrGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 333
QY 944 TTGTGTGACCAAGATATATATATATATATATATATATATATATATATATATATATATAT 1003
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 334 TrpLysAsnLysGlyTyrIleLeuMetAlaTyrAsnLysAsnAlaCysGlyIleAla 353
QY 1004 AGTTATCCCTCTTACCCCAAAATC 1027
  
```

DB 354 AsnLeuAlaSerPheProLysMet 361

RESULT 9

AAM7071 standard; Protein: 329 AA.

AAM7071;

09-NOV-1998 (first entry)

Rat Cathepsin K polypeptide.

Rat; cathepsin K; cysteine protease; osteoclast; bone resorption;

osteoporosis; periodontal disease; arthritis; cancer; tumour metastasis;

Paget's disease.

Rattus sp.

EP861898-A1.

02-SEP-1998.

23-DEC-1997; 97EP-0310523.

26-FEB-1997; 97US-0806959.

(SMK) SMITHKLINE BEECHAM CORP.

Brun K, Field JA;

WPI; 1998-449110/39.

N-PSDB; AAV48225.

DNA encoding rat cathepsin K polypeptide - and corresponding e.g.

polypeptide, antibody, agonist and antagonist etc.

Claim 15; Fig 1; 19pp; English.

The Rat cathepsin K is a novel cysteine protease found to be abundantly

and selectively expressed in osteoclasts, which suggests a role in bone

resorption. The cathepsin K polypeptides and polynucleotides can be used

to produce agonists and antagonists to treat e.g. osteoporosis,

periodontal disease, arthritis, cancer, tumour metastasis and Paget's

disease.

Sequence 329 AA:

Alignment Scores:

Prod. No.: 5,596-95

Score: 937.50

Percent Similarity: 70.73%

Best Local Similarity: 54.57%

Query Match: 46.95%

DB: 21 Gaps: 4

US-10-010-580-1 (1-1126) x AAM7071 (1-329)

```

QY 54 CTGCTCCCTGTCCTCTATGTCAGTGGCACAAGTACATTAAGATCCACTGTGATCAT 113
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 8 LeuLeuProValSerPheLea-----ProGluGluLeuLeuSptThr 24
QY 114 CACTGCAATCTGCGAAGAAACCTACAGCAAAACATCAAGAGACAGAT--GAGGAAGTA 172
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 25 GlnTyrGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 44
QY 173 GCAAGGCGTCTATCTGGAAAAAAATTTAAATTTGTGATGCTTCAATCTGGAAACAC 232
    .....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 45 SerArgArgLeuLeuIleProGluLysAsnLeuLysLysLysLysLysLysLysLysLys 64
QY 233 TCAAGATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 292
    |||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
DB 65 SerLeuGlyAlaHisThrTyrGlnLeuAlaMetAsnHisLeuGlyAspMetThrSerGlu 84
  
```


[illegible]

PS Disclosure: Fig 2: 34pp: English

xx The invention relates to novel human osteoclast-specific or -related cDNA sequences. The sequence represents human cathepsin X. The sequences may have a use in gene therapy. The sequences of the invention are useful in the production of gene products useful in the therapeutic treatment or diagnosis of disorders involving aberrant bone resorption, for generating peptides which are useful for producing antibodies for identifying osteoclast-specific or -related peptides or gene products.

xx Sequence 329 AA:

Alignment Scores:

Align. No.:	2, 01e-94	Length:	329
Score:	935.50	Matches:	182
Percent Similarity:	70.12%	Conservative:	48
Best Local Similarity:	55.49%	Mismatches:	89
Query Match:	46,70%	Indels:	10
DB:	23	Gaps:	4

US-10-010-580-1 (1-1126) x ABB81810 (1-329)

QY CTCTCTCCCTTGGCTCCATGATGAGACAAAGTATTAAGATCCCATCTGATCAT 113

Db LeuNeuProVaLaSerHea-----LeuTYrProGluGluLeuAspThr 24

114 CACTGCATCTCTCGAAGAAACCTACAGCAACAACTAACAGAGAACT-GAGCAAGTA 172

25 HistProLeuNeuTrpLysThrHisArgLysGlnGlyAsnGlyValAspIle 44

QY 173 GCACGGCGTCTCATCTCGCAAAAAATCTAAAAATTGTCATGCTTCACATCTGGAAC 222

Db SerProArgLeuLeuTrpLysAsnLeuLysTyrIleSerIleHisAsnLeuGluAla 64

QY 233 TCATGTGAATGTCATTCATATATATATAGCATGACACATCTGGAGAGATGATGCTGAA 292

Db SerLeuGluAlaHisThrTrpGluLeuAlaMetAsnHisLeuGlyAspMetThrSerGlu 84

QY 293 GAACTGATATCTTTGATGGGTTCTCGAGAGTTCTCCAGCCAAATGCGAGACA-----AAT 346

Db GluValValAlaIleLysMetThrGlyLeuLysValProLeuSerHisSerArgSerAsp 104

QY 347 GTGCCTAATGAGCAACACTAATGCAAGATGTCGATCTCTGTCGACCTGCGAGAGAC 406

Db ThrLeuTrpIleLeuProGluTrpGluGlyAlaGlnAlaProAspSerValAspTyrArgLysLys 124

QY 407 GACTGTGTTACTGAAGTGAATACTAACTTCTTCTGCTGCTTCTTCTTCTTCTTCTTCT 466

Db GLyValValThrProValLysAsnGlnGlyCnGlySerGlySerGlyTrpAlaIlePheSer 144

QY 467 GTGGGGCCCTGGAACACACACTGAAAGCTAAAGCAAGAAAGCTGGTCTCTGATGCA 526

Db ValIleValLeuGluLeuGlnLeuLysLysLysThrGlyLysLeuLeuAsnLeuSerPro 164

QY 527 CAGAACTTGGTAGATTGCTCACTGAAAAAATATGCGAATAAGCGCTGCACATGCCGCTTC 586

Db GluAsnLeuValAspCysValSerGlu-----AsnAspGlyCysGlyGlyGlyTyr 184

QY 587 ATGACAACTGCTTTCTGATATATATATATATGATAACTAACTGATTTCTTCTTCTAT 646

Db MetThrAsnAlaIlePheGlnTrpValCnLysAsnArgGlyLysIleAspSerCnLysAspAlaTyr 201

QY 647 CTCTATCAAAACATCATGAAATGAGAAAGTGCAGATATTAATCAAAAAACCTGCTGCTG 706

Db ArgGlyLysTrpArgGluIleProGlnGlyAsnGlnLysAlaLeuLysArgAlaValAlaArg 241

QY 707 TCAAAATATACCTGAACCTCTCTTCTGCGAGTGAAGATGCTTAAAGAAAGCTGTGCGCAAT 766

Db ValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheGlnPheTyrSerLys 261

Db 242 ValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheGlnPheTyrSerLys 261
 Oy 827 GGTGTTACTATGAAATCTCTGT--ACTGAAATGTAATCATGAGTATTAGTGGTT 883
 Db 262 GlyValIleTyrAspLeuSerLysAsnSerAspAsnLeuAsnHisAlaValLeuAlaVal 281
 Oy 884 GCGTATGTAACCTTAATGCGAAACACTACTGCTTGTCGAAAACAGCTGGCCCTGCAC 943
 Db 282 GlyTyrGlyIleGlnLysGlyAsnLysHisIleLeuLysAsnSerTyrPGLysLysn 301
 Oy 944 TTGGGACCGAAGATATATTCGATGCGAAGAACAGTGAATCTGTTGGATTGCT 1003
 Db 302 TrpGlyAsnLysGlyIleLeuMetAlaIleArgAsnLysAsnAlaIleGlyIleAla 321
 Oy 1004 AGTTATCCCTTACCCGAAATC 1027
 Db 322 AsnLeuAlaSerPheProLysMet 329
 RESULT 13
 ID AAM48939 standard: Protein: 329 AA.
 AC AAM48939:
 DE 26- OCT-1998 (first entry)
 XX Mutant human cathepsin K C139A protease.
 DE Mutant human cathepsin K C139A protease.
 KW Mutant human cathepsin K protease; SPA: tyrosine phosphatase;
 KW scintillation proximity binding assay; diabetes; cancer;
 KW 4-phosphono(difluoromethyl) phenylalanine group; osteoporosis.
 OS Homo sapiens.
 OS Synthetic.
 FH Key location/Qualifiers
 FT Peptide 1..15 /note= "Pre-domain"
 FT Peptide 16..114 /note= "Pro-domain"
 FT Protein 115..329 /note= "Human cathepsin K C139A"
 FT Misc-difference 139 /note= "changed from Cys in wild-type to Ala in
 mutant."
 FT
 FT
 FT
 PN W09820024-A1.
 XX 14-MAY-1998.
 XX 03-NOV-1997; 97WO-CA00824.
 XX 04-MAY-1996; 96US-0030411.
 XX (MERK) MERCK FROST CANADA INC.
 XX Desmarais S, Friesen R, Zamboni R;
 XX WPI: 1998-348101/30.
 DR N-PSDH; AAV32610.
 XX
 XX Peptide(s) useful in binding assays for tyrosine phosphatases or
 PT cysteine proteases - contain two or more 4-phosphono(difluoromethyl)
 PT phenylalanine groups to improve binding affinity
 PT
 XX
 XX
 PS Disclosure: Pages 43-44: 59pp: English.
 CC The present sequence represents a mutated human cathepsin K C139A
 CC protease. The invention provides a method for use in a scintillation
 CC proximity binding assay (SPA) for proteases and phosphatases. The
 CC method involves using mutated proteases and phosphatases whereby the
 CC catalytic cysteine residue of the enzymes are replaced with a serine or
 CC alanine residue to correct the problem of interference in SPA from

CC extraneous oxidizing and alkylating agents. The mutation affects the
 CC catalytic properties of the enzyme but does not affect their binding
 CC properties. The invention claims for new ligands for use in SPA which
 CC have increased binding affinity for a tyrosine phosphatase or cysteine
 CC protease. The ligands contain at least two 4-phosphono(difluoromethyl)
 CC phenylalanine groups which increase binding affinity of the ligand to
 CC its respective enzyme. The assay can be used to determine the ability
 CC of new ligands and compound mixtures to competitively bind with an
 CC enzyme. The method is claimed to allow a better usage of SPA in the
 CC discovery of compounds for the treatment and study of diseases, e.g.
 CC diabetes, cancer and osteoporosis.

CC Sequence 329 AA:

CC Alignment Scores:

Pred. NO.: 3,35e-94 Length: 329
 Score: 930.50 Matches: 182
 Percent Similarity: 70.128 Conservative: 48
 Best Local Similarity: 55.498 Mismatches: 89
 Query Match: 46.598 Indels: 10
 DB: 19 Gaps: 4

US-10-010-580-1 (1-1126) x AAM48939 (1-329)

Oy 54 CTGCTTCGCTTGTGCTCTGATGCGACGAAAGTACATAAAGATCCACTGCGATCAT 113
 Db 8 LeuLeuProValValSerPheAla-----LeuTyrProGlnGluIleLeuAspThr 24
 Oy 114 CACTGCAATCTCTGGAAGAAACCTACAGCAACCAATCAAGAGAGAT-GAGGAGATA 172
 Db 25 HistProLeuLeuTyrPylsLysThrHisATGGLysGlnTyrAsnAsnLysValAspGluLe 44
 Oy 173 GCACGGCGTCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCACAAATGCTGAAAC 232
 Db 45 SerArgArgLeuLeuTyrPylsLysAsnLeuLysTyrIleSerIleHisAsnLeuGluAla 64
 Oy 233 TCAATGGCAATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 292
 Db 65 SerLeuGlyValHisThrTyrGlnLeuAlaMetAsnHisLeuGlyLysPheThrSerGlu 84
 Oy 293 GAAGTATATCTTGTATGGTCCCTGAGAGTCCCAAGCCATAGCAGAGA-----AAT 346
 Db 85 GluValValGlnLysMetThrGlyLeuLysValProLeuSerHisSerArgSerAsnAsp 104
 Oy 347 GTCACTTATAGCTGAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 406
 Db 105 ThrLeuTyrIleProGluTyrPylsLysValArgAlaProAspSerValAspIleArgLys 124
 Oy 407 GCGTCTGTTACTGAAGTAAATACAGGGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 466
 Db 125 GlyTyrValThrProValLysAsnGlnGlyLysGlySerAlaIlePheSerSer 144
 Oy 467 GTGGGGGCGCTGGAAGCAGACAGCTGAGCTGAAACAGAGAAAGTGGTGTCTTGAAGTGA 526
 Db 145 ValGlyAlaLeuGlnGlyLysLeuLysLysLysThrGlyLysLeuAsnLeuSerPro 164
 Oy 527 CAGAACTGTAGATGCTCACTGAAATATATATATATATATATATATATATATATATATAT 586
 Db 165 GlnAsnLeuValAspCysValSerGlu-----AsnAspGlyCysGlyGlyGlyTyr 181
 Oy 587 ATGACAACTGCTTTCGAT 646
 Db 182 MetTrpAsnAlaPheGlnTyrValGlnLysAsnArgIleLeuAspSerGlnAspAlaTyr 201
 Oy 647 CCTTCAAAAGCCATGATGAGAAAGTCAATATATATATATATATATATATATATATATATAT 706
 Db 202 ProTyrValGlyClnGlnGlnLysSerCysMetTyrAsnProThrGlyLysAlaIleCys 221
 Oy 707 TCAAACTATGAACTTCCCTTGGCAGTGAAGATGCTTAAAGAAAGTGTGATGAT 766
 Db 222 ArgGlyTyrArgGlnIleLeuProGlnLysGlnGlyLysAlaLeuLysValAlaValArg 241
 Oy 767 AAGGACCTGT 826


```
UY 767 AAAGACCTGTCGTCGCTATAGATGCGAGCCACATTCCTTCTGCTACAGAACT 826
    |||
DB 242 ValGlyProValSerValAlaIleAspAlaSerLeuThrSerPheGlnPheTyrSerLys 261
    |||
UY 827 GATGCTTACTATGCAACCAATCCCTGTT--ACTGCAAGATGTGAATGCAGATATTAGTGGTT 883
    |||
DB 262 GlyValTyrTyrAspCysSerCysAsnSerAspAsnLeuAsnHisAlaValIleGluAlaVal 281
    |||
UY 884 GACTATGGTACGCTTAATGGCAAGATCTACTGTTGTGAAAGAAATAGTGGGCTGCAAC 943
    |||
DB 282 GlyTyrGlyIleGlnIleGlyAsnIleHisTyrPheIleIleLysAsnSerTTPGlyGluAsn 301
    |||
UY 944 TTGCGTCAGCAAGCATATATTCGATGCGCAAGAAACAGTGCAGATCACTGTGCGATTGCT 1003
    |||
DB 102 TrpGlyAsnLysGlyTyrIleLeuMetAlaIleAsnLysAsnAsnAlaCysGlyIleAla 321
    |||
UY 1004 ATTATTCGCTCTTACCCAGAAATC 1027
    |||
DB 122 AsnLeuAlaSerPheProLysMet 329
    |||
```

Search completed: June 27, 2003, 11:28:42
Job time : 82 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

```
Run on:      June 27, 2003, 11:23:59 ; Search time 49.5 Seconds
              (without alignments)
              4373.631 Million cell updates/sec
```

```

Title:      US-10-010-580-1
Perfect score: 1997
Sequence:   1 gcatattagtatgagca.....aataatgtgtcatgacct 1126

```

Scoring table:	
BLOSUM62	
Xgapop	10.0, Xgapext 0.5
Ygapop	10.0, Ygapext 0.5
Fgapop	6.0, Fgapext 7.0
Delop	6 0, Inlext 7 0

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 566448

```

Maximum seq length: 2000000000
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```

Command line parameters:
-MODEL=trame+nc2p.model -DEV=xlp
-O/-qgn2.1 /USPTO.spoof/US10010580/runat.27062003.104536.11748/app.query.fasta_1.1267
-DH=fr1.73 /OFTO.fstestn -SUP1X5+nc2p.rpr -MNMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blissum62 -TRANS=human40.cdf -LIST=45
-DOCLINK=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pico -OSB=ext -HEP5SIZE=500 -MATEXEN=2000000000
-USER=US10010580.qcgn.1.1.62.4runat.27062003.104536.11748 NCPD=6 -ICPU=3
-NO_MMAB -LARGESUBQUERY NEG.SCORES=0 -WAIT -DSPBLCK=100 -LONELOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAP=EXT=0.5 -FGAP=6
-FCGAP=EXT=7 -YGAP=0.1 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

```
Database : PIR_73.*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	156.2	78.2	331	2	A42482	cathepsin S (EC 3
2	134.4	66.8	310	2	A45087	cathepsin S (EC 3
3	108.9	54.5	217	2	S15844	cathepsin S (EC 3
4	94.4	47.3	324	2	A49868	cathepsin K (EC 3
5	93.9	47.0	329	2	JC2476	cathepsin K (EC 3
6	92.2	46.2	329	2	S74227	cathepsin K (EC 3
7	85.5	41.8	334	1	KHMSL	cathepsin L (EC 3
8	82.5	41.2	334	1	KHRTL	cathepsin L (EC 3
9	80.6	40.4	333	1	KHHUL	cathepsin L (EC 3
10	80.4	40.3	334	1	A58195	cathepsin L (EC 3
11	78.1	39.1	313	2	S47433	cathepsin L (EC 3
12	77.7	38.9	323	2	S19650	cathepsin L (EC 3
13	77.0	38.6	338	2	JC5443	cysteine proteinase
14	76.5	38.5	337	2	T24367	cathepsin L-like (probable cysteine

15	617.5	30.9	376	1	KHOP
14	618.5	31.0	364	2	T46630
43	619	31.0	376	2	EH5435
44	620	31.1	331	2	D86413
42	621.5	31.0	466	2	IO6416
41	623.5	31.2	364	2	T12039
39	633.5	31.6	427	2	S57776
38	637	31.9	493	2	T01206
37	638	31.9	455	2	T12041
36	639.5	32.0	367	2	T06529
35	644	32.2	454	2	KJ4848
34	647	32.4	471	1	KHRZ08
33	651.5	32.6	317	2	S44151
32	654	32.7	368	2	S47312
31	660.5	33.1	326	2	T09259
29	667	33.4	462	2	TN0719
28	675	33.8	458	1	KHRZ0A
27	684.5	34.3	218	2	S67481
26	687.5	34.9	218	1	KHCHL
25	706.5	35.4	480	2	T01207
24	709.5	35.5	326	2	S43991
23	715.5	35.8	322	2	S19649
22	724.5	36.3	331	2	JC5442
21	727.5	36.4	331	2	JC5441
20	734	36.8	343	2	UC7183
19	748	37.5	333	1	I52525
18	757	37.9	320	2	S19651
17	760.5	38.1	344	2	JX0366
16	760.5	38.1	339	2	A53027
15	764	38.3	326	2	S53027
14	765.5	38.3	339	2	A53810
13	765.5	38.1	344	2	JX0366
12	765.5	38.1	344	2	JX0366
11	765.5	38.1	344	2	JX0366
10	765.5	38.1	344	2	JX0366
9	765.5	38.1	344	2	JX0366
8	765.5	38.1	344	2	JX0366
7	765.5	38.1	344	2	JX0366
6	765.5	38.1	344	2	JX0366
5	765.5	38.1	344	2	JX0366
4	765.5	38.1	344	2	JX0366
3	765.5	38.1	344	2	JX0366
2	765.5	38.1	344	2	JX0366
1	765.5	38.1	344	2	JX0366

ALIGNMENTS

RESULT 1
A42482
cathepsin S (EC 3.4.22.27) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1999
C:Accession: A42482; A53625; A42896
R:Shi, G.P.; Mungert, J.S.; Meara, J.P.; Rich, D.H.; Chapman, H.A.
J. Biol. Chem. 267, 7258-7262, 1992
A:Title: Molecular cloning and expression of human alveolar macrophage cathepsin S, a
A:Reference number: A42482; MUID:92218373; PMID:1373132
A:Accession: A42482
A:Molecule type: mRNA
A:Residues: 1-331 <SH>
A:Cross-references: GB:93414; NID:q248405; PIDN:AAH22005.1; PID:g248406
A:Experimental source: alveolar macrophage
A:Note: sequence extracted from NCBI database (NCBIN:93414, NCBIPI:93443)
P:Shi, G.P.; Webb, A.C.; Foster, K.E.; Knoll, J.H.M.; Lemire, C.A.; Mungert, J.S.; Chan-
J. Biol. Chem. 269, 11530-11536, 1994
A:Title: Human cathepsin S: chromosome1 localization, gene structure, and tissue distri-
A:Reference number: A53625; MUID:94209337; PMID:8157683
A:Accession: A53625
A:Molecule type: DNA
A:Residues: 1-210, 'H', 212-331 <SH2>
A:Cross-references: GB:007374
R:Wiederanders, B.; Browne, D.; Kirschke, H.; von Figura, K.; Schmidt, R.; Peters, C.
J. Biol. Chem. 267, 13708-13713, 1992
A:Title: Phylogenetic conservation of cysteine proteolases. Cloning and expression of
A:Reference number: A42896; MUID:92317106; PMID:1377692
A:Accession: A42896
A>Status: not compared with conceptual translation
A:Molecule type: mRNA; Protein
A:Residues: 1-91, 'M', 93-160, 'S', 162-331 <WIE>
A:Cross-references: GB:S19127; GB:M90696; NID:q250802; PIDN:AAA09269.1; PID:g250803
A:Experimental source: testis
A:Note: sequence extracted from NCBI backbone (NCBIPI:107806)
C:Genetics:
A:Gene: GDB:CTSS
A:Cross-references: GDB:132414; OMIM:116845

Score: 944.50 Matches: 186
Percent Similarity: 69.55% Conservativity: 47
Best Local Similarity: 55.52% Mismatches: 79
Query Match: 47.30% Indels: 24
Indels: 2 Gaps: 5

US-10-010-580-1 (1-1126) x A49868 (1-329)

```

UY 54 CTGCTTCGGTGTGCTGCTATGTCAGTGCACATACATATAAGATCCGACTCTGGATCAT 113
    |||||
Db 8 LeuLeuProValValSerPheAla-----LeuHisProGluIuIleLeuAspThr 24
    |||||
UY 114 CACTGGAATCTCTGGAAGAAAACCTACAGCAACAACTCAAGAGAGAAAT-GAGGAAGTA 172
    |||||
Db 25 GlnTrpGluLeuTrpLysThrLysSerLysGlnTrpAsnSerLysValAspGluIle 44
    |||||
UY 173 GCAGCGGCTCTCATTTGAGAAAAAATTTGATGCTTCACAACTTCGACAC 232
    |||||
Db 45 SerTrpArgLeuIleTrpGluLysAsnLeuLysHisLysSerLysHisAsnLeuGluAla 64
    |||||
UY 243 TCAATGGGAATGCTATCATATGATCATGAGCATGAGCATGAGCATGAGCATGAGTA 292
    |||||
Db 65 SerLeuGluValHisThrLysGluLeuAlaMetAsnHisLeuGluLysAspMetThrSerGlu 84
    |||||
UY 293 CAACCTATATCTTCATGCTCCCTCAGACACTCCACCAATGCCAGACAATGTCAC 352
    |||||
Db 85 GluValValGlnLysMetThrGlyLeuLysValProProSer----- 98
    |||||
UY 353 TATAGTCAAACTCTATTCAGAAATG-----CGCAT 365
    |||||
Db 99 ArgSerHisSerAsnAspThrLeuLysIleProAspTrpGluGlyAsnThrProAsp 117
    |||||
UY 386 TTTGTGGAAGTGAAGAGAAATGCTGTTATGTAATGTAATGCAAGGCTTCCTGCT 445
    |||||
Db 118 SerLeuAspTrpAlaGlyLysGlyLysValThrProValLysAsnGlnGlyLysGly 137
    |||||
UY 446 GCTTGTGGCGCTTCACGCGCTGGCGGCTCGAAGCAACCAAGCAAGCAAGCAAGCA 505
    |||||
Db 138 SerGlySerTrpAlaPheSerSerValGlyAlaLeuGluIuLysLeuLysLysThrGly 157
    |||||
UY 506 AAGCTGCTGTCTGATGTCAGACAGACATGCTGATGCTGCAACTGCAAAATGCGAAT 565
    |||||
Db 158 LysLeuLeuAsnLeuSerProGlnAsnLeuValAspLysValSerGluAsnTrp----- 175
    |||||
UY 566 AAGATGTCATGTCGCTTATATACACATGCTTTCATGATATATATATATATATAT 625
    |||||
Db 176 GlyLysGluLysGlyLysGlyLysMetThrAsnAlaPheGlnTrpValGlnAsnAsnArgGly 194
    |||||
UY 626 ATTGATTCAGACACTTCCTATTCCTACCAAGCAATCAATGCAAGAGTCCACATATGCTCA 685
    |||||
Db 195 IleAspSerGluAspAlaTrpProTyrValGlyGlnAspGluSerCysMetTyrAsnPro 214
    |||||
UY 686 AAAAAAGATGTCATATATTTTAAATATATATGAAATTTTGTGTAAGTGAAGATGCC 745
    |||||
Db 215 ThrGlyLysAlaAlaLysCysAspArgGlyTyrArgGluIuIleProGluGluLysAsnGluLysAla 234
    |||||
UY 746 TTTAAAGAAATTTTGTGTAATTAAGAAATTTTGTGTAATTTTGTGTAATTTTGTGTA 805
    |||||
Db 235 LeuLysArgAlaValAlaAlaGlyValGlyProValSerValAlaIleAspAlaSerLeuThr 254
    |||||
UY 806 TTTCTTCTCCGTACAGACAGTGTGCTACTATGACCAATCCATCTGATCAT 862
    |||||
Db 255 SerPheGlnIlePheTrpSerLysGlyValLysLysLysLysLysLysLysLysLysLys 274
    |||||
UY 863 AATATATGAGATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 922
    |||||
Db 275 AsnHisAlaValIleValAlaGlyTyrGlyIleGlnLysGluLysAsnLysHisLysLysLys 294
    |||||
UY 923 AAAAAAGATGTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 982
    |||||
Db 295 LysAsnSerTrpAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 314
    |||||
UY 983 GCAAAATCATCTGCGAATTCCTACTTAATCCTCTTACCAGCAAAATC 1027
    |||||

```

Db 315 AsnAsnAlaCysGlyIleAlaAsnLeuAlaSerPheProLysMet 329

RESULT 5

JC2476

cathepsin K (EC 3.4.22.-) precursor - human

N/Alternate names: cathepsin O2

C/Species: Homo sapiens (man)

C/Date: 21-Feb-1995 #sequence, revision 05-Apr-1995 #text, change 22-Jun-1999

C/Accession: JC2476; S55763; S68459; I38752; S48830

R/Inoaka, T.; Bilbe, G.; Ishibashi, O.; Tezuka, K.; Kamegawa, M.; Kokubo, T.

Biochem. Biophys. Res. Commun. 206, 89-96, 1995

A/Title: Molecular cloning of human cDNA for cathepsin K: Novel cysteine proteinase p

A/Accession number: JC2476; M01D:95118380; P01U:7818555

A/Accession: JC2476

A/Molecule type: mRNA

A/Residues: 1-329 <TNA>

A/Cross-references: EMBL:X82153; NID:g562756; P01U:CAA57649.1; P1D:g562757

R/Broemme, D.; Okamoto, K.

Biol. Chem. Hoppe-Seyler 376, 379-384, 1995

A/Title: Human cathepsin O2, a novel cysteine proteinase highly expressed in osteoclast

A/Reference number: S55763; M01D:96082523; P01U:7576232

A/Accession: S55763

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-329 <SBRO>

A/Cross-references: GB:S57895; NID:g119555; P01U:AA835521.1; P1D:g1195556

R/Shi, G.P.; Chapman, H.A.; Bharti, S.M.; Deleuw, C.; Reddy, V.V.; Weiss, S.J.

FEBS Lett. 357, 129-134, 1995

A/Title: Molecular cloning of human cathepsin O, a novel endoprotease and homologue

A/Reference number: I38752; M01D:95104457; P01U:7805878

A/Accession: S68459

A/Molecule type: mRNA

A/Cross-references: EMBL:M13665; NID:g606922; P01U:AAA65233.1; P1D:g606923

A/Experimental source: tissue-type blood

C/Genetics:

A/Genes: GDB:CTSK; PYCD; CTSO1; CTSO2; PKND; CTSO

A/Cross-references: GDB:453910; OMIM:265800; OMIM:600550; OMIM:601105

A/Map position: 1q21-1q21

C/Superfamily: papain

C/Keywords: cysteine proteinase; glycoprotein; hydrolase

F1-15/Domain: signal sequence; status predicted <PRO>

F16-114/Domain: signal sequence; status predicted <PRO>

F115-329/Product: cathepsin K status predicted <MT>

F103,161,213/Binding site: carbohydrate (Asn) (covalent) *status predicted

F139,276,296/Active site: Cys, His, Asn *status predicted

DB:

Alignment Scores:			
Pred. No.:	9,15e-74	Length:	329
Score:	939.50	Matches:	183
Percent Similarity:	70.43%	Conservative:	48
Best Local Similarity:	55.79%	Mismatches:	88
Query Match:	47.05%	Indels:	10
DB:	2	Gaps:	4

US-10-010-580-1 (1-1126) x JC2476 (1-329)

```

UY 54 CTGCTTCGGTGTGCTGCTATGTCAGTGCACATACATATAAGATCCGACTCTGGATCAT 113
    |||||
Db 8 LeuLeuProValValSerPheAla-----LeuTrpProGluIuIleLeuAspThr 24
    |||||
UY 114 CACTGGAATCTCTGGAAGAAAACCTACAGCAACAACTCAAGAGAGAAAT-GAGGAAGTA 172
    |||||
Db 25 GlnTrpGluLeuTrpLysThrLysSerLysGlnTrpAsnSerLysValAspGluIle 44
    |||||
UY 173 GCAGCGGCTCTCATTTGAGAAAAAATTTGATGCTTCACAACTTCGACAC 232
    |||||
Db 45 SerTrpArgLeuIleTrpGluLysAsnLeuLysHisLysSerLysHisAsnLeuGluAla 64
    |||||
UY 243 TCAATGGGAATGCTATCATATGATCATGAGCATGAGCATGAGCATGAGCATGAGTA 292
    |||||
Db 65 SerLeuGluValHisThrLysGluLeuAlaMetAsnHisLeuGluLysAspMetThrSerGlu 84
    |||||

```


QY 927 ACACCTGGGCGCCACACTTTCGACCAAGCATATATTCGACATGCAACAAACAGTGGAA 986
 DB 300 snstftrpglylyglutfrplymetlasrplytrflllelyslleallalyaspargasm 320
 QY 987 ATACCTGCGATTCCTACCTTATCCCTTACCCA 1021
 DB 320 snstfscysglylualatfrAlaAlaIasertyrpro 331

RESULT 9

KHHHL
 N: Alternate names: precursor [validated] - human
 C: Species: Homo sapiens (man)
 C: Date: 30-Sep-1990 #sequence_revision 30-Sep-1990 #text_change 08-Dec-2000
 C: Accession: S01002; B12333; S09065; A45043; S00323; B27011; A26069; A32683; D27011; F27
 R: Gal, S.; Gottesman, M.M.
 Biochem. J. 253, 303-306, 1988
 A: Title: Isolation and sequence of a cDNA for human pro-(cathepsin L).
 A: Reference number: S01002; MUID:88339905; PMID:3421948
 A: Accession: S01002
 A: Molecule type: mRNA
 A: Residues: 1-333 <GAD>
 A: Cross-references: GH:K12451; NID:q29714; PIDN:CAA30981.1; PID:q29715
 R: Joseph, L.; Chang, L.C.; Stamenkovich, D.; Sukhame, V.P.
 J. Clin. Invest. 81, 1621-1629, 1988
 A: Title: Complete nucleotide and deduced amino acid sequences of human and murine prepro
 A: Reference number: A92768; MUID:88213715; PMID:2835338
 A: Accession: B12333
 A: Molecule type: mRNA
 A: Residues: 1-333 <JOS>
 A: Cross-references: GH:M20496; NID:q809235; PIDN:AAA66974.1; PID:q190418
 R: Joseph, L.; Lapid, S.; Sukhame, V.
 Nucleic Acids Res. 15, 3186, 1987
 A: Title: The major ras induced protein in NIH3T3 cells is cathepsin L.
 A: Reference number: S09065; MUID:87174843; PMID:3550705
 A: Accession: S09065
 A: Molecule type: mRNA
 A: Residues: 11-154 <JOS>
 A: Cross-references: EMBL:X05256; NID:q29718; PIDN:CAA28877.1; PID:q1340178
 R: Chaudh, S.S.; Popescu, N.C.; Ray, D.; Fleischmann, R.; Gottesman, M.M.; Troen, B.R.
 J. Biol. Chem. 268, 1039-1045, 1993
 A: Title: Cloning, genomic organization, and chromosomal localization of human cathepsin
 A: Reference number: A45043; MUID:9121212; PMID:8419312
 A: Accession: A45043
 A: Molecule type: DNA
 A: Residues: 40-46;82-86;130-135;205-210;259-264;299-304 <CHA>
 A: Cross-references: GH:I06426
 A: Note: only exon-intron splice junctions are shown
 R: Ritonja, A.; Popovic, T.; Kotnik, M.; Machleidt, W.; Turk, V.
 FEBS Lett. 228, 141-145, 1988
 A: Title: Amino acid sequences of the human kidney cathepsins H and L.
 A: Reference number: S00322; MUID:88137635; PMID:3342889
 A: Accession: S00322
 A: Molecule type: Protein
 A: Residues: 'X', 115-129, 'M', 131-133, 'E', 135-141;292-307, 'TD', 310-333 <MA2>
 R: Machleidt, W.; Ritonja, A.; Popovic, T.; Kotnik, M.; Brzin, V.; Machleidt, T.
 In Cysteine Proteinases and Their Inhibitors, Turk, V., ed., pp.3-18, Walter de Gruyter,
 A: Title: Human cathepsins B, H and L: characterization by amino acid sequences and some
 A: Reference number: A27011
 A: Accession: A27011
 A: Molecule type: Protein
 A: Residues: 'X', 115-129, 'M', 131-133, 'E', 135-141;292-307, 'TD', 310-333 <MA2>
 R: Mason, R.W.; Walker, J.E.; Northrop, F.D.
 Biochem. J. 240, 373-377, 1986
 A: Title: The N terminal amino acid sequences of the heavy and light chains of human cathe
 A: Reference number: A26069; MUID:87127952; PMID:3545185
 A: Accession: A26069
 A: Molecule type: Protein
 A: Residues: 114-147, 'P', 149-152, 'Y', 292-333 <MAS>
 R: Smith, S.M.; Gottesman, M.M.
 J. Biol. Chem. 264, 20487-20495, 1989
 A: Title: Activity and deletion analysis of recombinant human cathepsin L expressed in Es

A: Reference number: A32683; MUID:90062183; PMID:2684978
 A: Contents: annotation
 C: Genes: GDB:CTSL
 A: Gene: GDB:CTSL
 A: Cross-references: GDB:119824, OMIM:116880
 A: Map position: 9q22.1-9q22.2
 A: Introns: 42/3; 83/3; 132/3; 207/3; 262/1; 301/2
 C: Complex: heterodimer of disulfide linked chains produced from a single chain precu
 C: Function:
 A: Description: catalyzes hydrolysis of peptide bonds in proteins
 A: Pathway: Intracellular protein degradation
 A: Note: Important role in the lysosomal degradation of proteins
 C: Superfamily: papain
 C: Keywords: cysteine proteinase; glycoprotein; heterodimer; hydrolase; lysosome; prot
 F:1-17/Domain: signal sequence #status predicted <STG>
 F:18-113/Domain: signal sequence #status predicted <PRO>
 F:114-333/Product: cathepsin L #status experimental <MAT>
 F:114-288/Product: cathepsin L heavy chain #status experimental <HCH>
 F:292-333/Product: cathepsin L light chain #status experimental <LCH>
 F:135-178,169-211,269-322/Disulfide bonds: #status predicted
 F:138,276,300/Active site: Cys, His, Asn #status predicted
 F:221/Binding site: carbonylate (Asn) (covalent) #status predicted

Alignment Scores:
 Pred. No.: 3,986-62 Length: 333
 Score: 806.00 Matches: 163
 Percent Similarity: 64.74% Conservative: 50
 Best Local Similarity: 45.54% Mismatches: 96
 Query Match: 40.36% Indels: 20
 DB: 1 Gaps: 7

US-10-010-580-1 (1-1126) x KHHUL (1-333)

QY 75 GCAGTGGCAACAAGTACATCAATGATCCACTGATCACTGATGATCTTGGAGAAA 134
 DB 15 AlasecAlatfrleuthrphesphsrlserleuGluAlaGlntrPrthlysrlypysala 34
 QY 135 ACCTACGACAAACATCAAG 194
 DB 35 MelHisasnArq--LeutyrglymetasnGluGluGlyTrpArqArqAlaValtrpGluL 54
 QY 195 AAAATGTAATTTTGATGATCTTCACATTCGACACATCAATGGCAATGCAATGATATG 254
 DB 54 yAsnMetlysmetlleGluLeuHisasnGlnGluItyrArgGluItylyshsSerPhe 74
 QY 255 ATCTAGCATGACCATCTGGAGACATGACTGTAAGAAATATATCTTGATGGTT 314
 DB 74 hMetAlaMetasnAlaPheGlyAspMetThrSerGluIthrArgGlnValMetasn 94
 QY 315 CCCTG-----ACAGTTCGACGCCA-----TGCCAGAGAAATGCTACTTACT 359
 DB 94 lypheGlnasnArqlysrProArqlysglyLyAsvalPheGlnGluProleupheryGluA 114
 QY 360 CAAACGTATATCAGAAATGGCTGATTCGTGAGTGGAGAGAGAGAGAGAGAGAGAGAG 419
 DB 114 lA-----ProArgserValAspTrpArgGluItyrValThrP 128
 QY 420 AAGTGAATACCAAGGTTTGT 479
 DB 128 roVallysasncInglylncysGlyserCysTrpAlaPheSerValaThrGlyAlaLeuc 148
 QY 480 AACGACACTGACAGCAAAACAGCAAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 539
 DB 148 lGlyGlnMetPheArqlysrThrArgAlaGlnleuSerleuSerleuGlnleuValA 168
 QY 540 ATTGCTCAACTGAAATATATGGAATTAAGCTGCAATGCGCTTTCATGCAACTGCTG 599
 DB 168 spCysserGlyProGln--GlyAsnGluGlyCysAsnGlyCylLeuMetAspTrpAlaP 187
 QY 600 TCCAGATATTTTATATTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 659
 DB 187 heGlnItyrValGlnAsnGlylncysGlyleuAspSerleuGlnleuSerlyrProtyrGlnAla 207

C:Accession: S47433
 R:Le Boulay, C.; van Wormhout, A.; Sellios, D.
 submitted to the EMBL Data Library, August 1994
 A:Description: Molecular cloning and sequencing of the two cDNAs that encode cathepsin L
 A:Reference number: S47432
 A:Accession: S47433
 A:Molecule type: mRNA
 A:Residues: 1-113 <1kB>
 A:Cross-references: EMBL:X80900; NID:q510735; PDB:CAA56915.1; PDB:q530736
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:122/260/280/Active site: Cys, His, Asn *status predicted

Alignment Scores:
 Pred. No.: 5,97e-60 Length: 313
 Score: 781.00 Matches: 158
 Percent Similarity: 65.53% Conservative: 53
 Host Local Similarity: 49.07% Mismatches: 101
 Query Match: 39.11% Indels: 11
 DB: Caps: 5

US-10-010 580-1 (1-1126) x S47433 (1-313)

```

UY      66  TCCCTCTATGCAGCTGGCAGCAATACATTAAGATCCGACTTCGATCATCATCGCAATCTC 125
DB      1  CysGlyLeuAlaLeuAla-----ThrAlaSerProSerTrrpLuhis 14
UY      126 TGGAGAAAGAACTTAAATTTTTCATCTTCACATCTGGAGACTCAATGGCAATC 244
DB      15  PheLysThrGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 34
UY      185 ATCTGGCAAAAAAATCTAAATTTTTCATCTTCACATCTGGAGACTCAATGGCAATC 244
DB      45  ValPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 54
UY      245 CATTCATATCATCTGCAGCATCAACCATCTGCAGCATCAACCATCTGCAGCATCAATCT 304
DB      55  ValThrPheGlyValAlaMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 74
UY      305 TTGATGATCTTCTTAAATTTTTCATCTTCACATCTGGAGACTCAATGGCAATC 364
DB      75  ValMetLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 94
UY      365 TCTAATACAGCAATTCCTGATCTGCTGCAGCATCAACCATCTGCAGCATCAATCT 424
DB      95  GlnGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 113
UY      425 AAATACGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 484
DB      114  LysAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 133
UY      485 CAATCTGAAAGTAAAGAAAGAAAGTGGTGTCTGAGTGCAGCAACTTGGTAGATTGC 544
DB      144  GlnHisPheLeuLysAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 153
UY      545 TTAACCTGAAAAAATGCGCAATTAAGCGCTGCAATGCGGCTTCATGACAACTGCTTCAC 604
DB      154  SerThrGln-----TyrGlnAsnAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 172
UY      605 TATATTTATGTAAGCAACGCAATGTAAGCAACGCTTCATGCTTCAGCAACGCAATG 664
DB      173  TyrIleLysAspAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 192
UY      665 GAAAGCTGAGATGATCTGAAAAAAGTGGTGTCTGAGTGCAGCAACTTGGTAGATTGC 724
DB      193  ArgSerCysArgPheAspAlaAsnSerIleGlnAlaThrCysThrGlyPheValGlnVal 212
UY      725 CCGCTTGGCAGTCAAGATGCTTAAAGCAACGCTGCGCCCAATTAAGCAACGCTGCTG 784
DB      213  GlnHis-----ThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 231
UY      785 GTTATAGATGTAAGCAACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 844
DB      111  LysThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln
  
```

DB 232 AlaIleAspAlaSerHisPheSerPheGlnPheTyrSerSerGlyValTyrTrpGlnLys 251
 OY 845 TCCGTGACT---CAGATGTGATCATGAGATATTAGTGGCTATGCTAAGCTTAAT 901
 DB 252 LysCysSerProThrAsnLeuAspHisGlyValLeuAlaValGlyTrpGlnThrGlnSer 271
 OY 902 CGGAACAGTACTGCTGCTGCAAAACAGCTGGGCGCTCAACTTGGTGCAGCAAGATAT 961
 DB 272 ThrLysPyrTrpPheValLysAsnSerTrrpLysSerGlyTrpPheLysPheValGly 291
 OY 962 ATTGGATGCGAAGAAACAGTGCAGATATCTGTTGAGTGTCTATGTTATCCCTTACCA 1021
 DB 292 IleLysMetSerArgAsnArgAspAsnAsnGlyGlyIleAlaSerGlnProSerTrrp 311
 OY 1022 GAAATC 1027
 DB 312 ThrVal 313

RESULT 12
 S19650
 cysteine proteinase (EC 3.4.22.-) precursor (clone LCP2) - American lobster
 C:Species: Homarus americanus (American lobster)
 C:Date: 04-Dec-1992 *sequence_revision 04-Dec-1992 *text_change 22-Jun-1999
 C:Accession: S19650; S31655
 R:Laycock, M.V.; Mackay, R.M.; Di Fruscio, M.; Gallant, J.W.
 FEBS Lett. 301, 125, 1992
 A:Title: Molecular cloning of three cDNAs that encode cysteine proteinases in the dlp
 A:Reference number: S19649; MUID:92070467; PMID:1995950
 A:Accession: S19650
 A:Molecule type: mRNA
 A:Residues: 1-323 <1kB>
 A:Cross-references: EMBL:X63568; NID:q11052; PDB:CAA5148.1; PDB:q11053
 R:Laycock, M.V.; Mackay, R.M.; Di Fruscio, M.; Gallant, J.W.
 FEBS Lett. 301, 125, 1992
 A:Title: Correction. Molecular cloning of three cDNAs that encode cysteine proteinase
 A:Reference number: S31654; MUID:93083613; PMID:1451782
 A:Accession: S31655
 A:Molecule type: mRNA
 A:Residues: 1-323 <1kB>
 A:Cross-references: EMBL:X63568; NID:q11052; PDB:CAA5148.1; PDB:q11053
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; disulfide bond; hydrolase; zymogen
 F:117-107/Domain: signal sequence *status predicted <Sig>
 F:117-107/Domain: propeptide *status predicted <Pro>
 F:108-323/Product: cysteine proteinase *status predicted <Mat>
 F:128-171,162-204,263-312/Disulfide bonds: *status predicted
 F:131,270/290/Active site: Cys, His, Asn *status predicted

Alignment Scores:
 Pred. No.: 1,34e-59 Length: 323
 Score: 777.00 Matches: 161
 Percent Similarity: 65.44% Conservative: 53
 Best Local Similarity: 49.24% Mismatches: 103
 Query Match: 38.91% Indels: 11
 DB: Caps: 6

US-10-010-580-1 (1-1126) x S19650 (1-323)

```

OY      48  GTTGGCTGCTTCCGTGCTGCTCTCTATGCAAGTGGCAACATGATTAAGATCCGATCTG 107
DB      3  ValAlaValAlaLeuPheLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 19
OY      108  GATATATCATGGAATCTCTGGAAGAAAGAACTACAGCAACATC-AAGCAAGAAATGAG 166
DB      20  -----TrrpLuhisPheLysGlyLysTrpGlnGlnGlnGlnGlnGlnGlnGlnGln 36
OY      167  GAACTACAGCGGCTGCTGCAATCTGCAAAAGAAAGAAAGTGGTGTCTGAGTGCAGCA 226
DB      37  AspSerTrpArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 56
OY      227  GAACATCTCAATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 286
DB      57  LysTrpGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 76
  
```

OY 287 GGTGAAGAGTGCATATCTTTCATGCGTTCCTGAGAGTCCCAACCAATGGCAGAGA--- 343
 |||||
 Db 77 LeuLIGLupheasnaAlaValMetLysGly---AsnIleProArgaSerAlaIproVal 95
 344 AATGCATCTTATAGCTCAAACTCAATCAGAAATTCCTGATCTTGCGATCGAGAGAG 403
 |||||
 OY 96 SerValPheTyrProLysLysGlyIuThrClyProGlnAlaThrGluValAspTPrArgThr 115
 |||||
 OY 404 AAGGCTGTGTGTAAGTAAATACCAGGGTCTTGCTGCTTGTGGCTTTGACG 463
 |||||
 Db 116 LysGlyAlaValAlaThrProValLysAspGlnGlyLncysGlySerCysTPrAlaPheSer 135
 464 GCTGCGGGCGCCGCGAAGCAGACACTGAAGCTAAAAAGCAAGACGCTGCTGAGT 523
 |||||
 Db 136 ThrThrGlySerLeuGlyGlyGlnHisPheLeuYThrGlySerLeuIleSerLeuAla 155
 524 GCACAGAACTTGATGATTCGTCACACAGTGAATAATAGGAATGAAGCGTCAATGGCGT 583
 |||||
 Db 156 GluGlnGlnLeuValAspCysSer---ArgProTyrGlyProGlnGlyCysAsnGlyOly 174
 584 TTCATGACAACTGCTTCCATATATATATGATAACAGCGCATTTGATCGAAGCTTCC 643
 |||||
 OY 175 TrpMetAsnAspAlaPheAspTyrIleLysAlaAsnAsnGlyIleAspThrGluAlaAla 194
 644 TATCCCTACAAAGCCATGCAATGGAAGTGCAGATATGACTGCAAAAAACCGAGCTGCCACA 703
 |||||
 Db 195 TyrProTyrGluAlaAlaArgAspOlySerCysAlaPheAspSerAsnSerValAlaAlaThr 214
 704 TGTTAAAGTATACAGAACTTCCTTTGGCAGTGAAGTGCCTTAAAGACGCTGGCC 763
 |||||
 OY 215 CysSerGlyHisIstHisnIleAlaSerGlySerGlyIuThrGlyLncGlnAlaValAlaArg 234
 764 AATAAGGACGCTGCTGCTGCTATAGATGAGAGCCACTATTTCTTCTGCTGACAGA 823
 |||||
 Db 235 AspIleGlyProIleSerValThrIleAspAlaAlaHisSerSerPheGlnPheYSer 254
 824 AGTGTGTCTACTATGAACCATCCCTGACTCAGAAAT---GTCAATCATGCACTATTAGT 880
 |||||
 OY 255 SerGlyValIlyTyrGlyIuProSerCysSerProSerTyrLeuAspHisAlaValLeuAla 274
 881 GTTGCTATGCTTAACCTTAATGCAAGACTGCTGCTGCAAAACAGCTGGCGCTC 940
 |||||
 Db 275 ValGlyTyrGlySerGlyGlyGlnGlyLncHisPheThrPleuValLysAsnSerTPrAlaThr 294
 941 AACTTGTGACCAAGATATATTCGATGGCAGAGAAACAGTGAATCACTGTGGGATT 1000
 |||||
 Db 295 SerTPrGlyAspAlaGlyTyrIleLysMetSerArgAsnArgAsnAsnAsnCysGlyIle 314
 1001 GCTACTTATCCCTCTTACCCA 1021
 |||||
 OY 315 AlaThrValAlaIleSerTyrPro 321
 |||||
 Db 315 AlaThrValAlaIleSerTyrPro 321
 |||||
 RESULT 13
 JCS443
 cathepsin L-like cysteine proteinase (EC 3.4.22.-) cl [similarity] - Maize weevil
 C:Species: Sitophilus zeamais (maize weevil)
 C:Date: 17-Jun-1997 #sequence: revision 18-Jul-1997 #text_change 20-Jun-2000
 C:Accession: JCS443
 C:Matsumoto, I.; Emori, Y.; Abe, K.; Arai, S.
 J. Biochem. 121, 464-476, 1997
 A:Title: Characterization of a gene family encoding cysteine proteinases of Sitophilus z
 nd germ cells.
 A:Reference number: JCS441: MUID:97279098: PMID:9133615
 A:Accession: JCS443
 A:Molecule type: mRNA
 A:Residues: 1-338 <MA>
 A:Cross-references: GB:082884; NID:92804261; PIDN:BAA24442.1; PID:92804262
 C:Comment: This enzyme acts as a secreted or lysosomal proteinase. It has various functi
 y organs, and cleaning in the malpighian tubule.
 C:Superfamily: papain
 C:Keywords: cysteine proteinase; hydrolase
 F:145,284,305/Active site: Cys, His, Asn *status predicted

Alignment Scores:
 Pred. No.: 4,95e-59 Length: 338
 Score: 770.50 Matches: 167
 Percent Similarity: 63.25% Conservative: 43
 Best Local Similarity: 50.30% Mismatches: 110
 Query Match: 38.58% Indels: 13
 DB: 2 Gaps: 5

US-10-010-580-1 (1-1126) x JCS443 (1-338)

OY 56 GCTTCGCTTCGCTCCATGACAGTGGCAGCAAGTACATAAAGATCCAGTTCGATCATCA 115
 |||||
 Db 8 AlaAlaValAlaIleSerCysGlnAlaValSer-PheTyrAsp---LeuValGlnGlnGly 26
 116 CTGGAATCTGTGGAGAAAGAACTCAGCAAGCAAAATC-AAGGAAGAGATGAGCAATGATC 174
 |||||
 OY 26 nTPPserSerPheLysMetGlnHisSerLysAsnTyrAspSerGlyIuThrGluIuArgH 46
 175 ACGCGCTCATCTGGGAAAAAATCTAAATTTGTGATGCTTCAACATCTGGACACCTC 234
 |||||
 Db 46 eArgMetLysIlePheMetGlnAsnAlaHisLysValAlaLysHisAsnLysLeuPhe 66
 235 AATGGGAATCATTCATATGATGATAGGCATGAACCATCTGGGACATGATGCTGTGACA 294
 |||||
 OY 66 rGlnGlyPheValLysPheLysLeuGlyLeuAsnLysTyrAlaAspMetLeuHisnIsc 86
 295 AGTGAATCTTGTGATG-----GGTCCCTGAGAGATTCCAGCCA 333
 |||||
 Db 86 urPheValSerThrLeuAsnGlyPheAsnLysThrLysAsnHisIleLeuLysGlySerAs 106
 334 ATGGCAGAGAAATGTCACTATTAGCTCAAACTCAATCAGAAATTCGCTGATCTGTGCA 393
 |||||
 Db 106 rPheAsnAspAlaValAlaArgPheIleSerProAlaAsnValLysLeuProAspThrValAs 126
 394 CTGAGAGAGAGAGCGCTGCTGCTTACTGAGTGAAGTGAATACAGAGTTCGCTGCTGTTCG 453
 |||||
 Db 126 pTPrArgAspLysGlyAlaValAlaThrGluValLysAspGlnGlyIleLysCysGlySer 146
 454 GGTTCAGCGCTGTGGGGCGCTGGAAATACAACTGAAGTAAACAGAAAGCTGGT 513
 |||||
 OY 146 pSerPheSerAlaThrClySerLeuGlnGlyGlnHisPheArgLysThrClyLysLeuVal 166
 514 GTCTGCTAGTGCACAGAACTTGATGATGCTGCTGCAATGGAATATATGCGAATTAAGCGT 573
 |||||
 Db 166 lSerLeuSerGlnGlnAsnLeuValAspCysSer---GlyArgTyrGlyLysAsnGlyC 185
 574 CAATGGCGCTTTCATGCAACCTGCTTCCAGTATATTATTGATTAACAACGCGATTCATT 633
 |||||
 Db 185 sAsnGlyOlyLeuMetAspAsnAlaPheArgTyrIleLysAspAsnGlyGlyIleAspThr 205
 634 AGAAGCTTCCTATCCCTACCAAGCCATGAAATGGAATGCAAGATATGACTCAAAAAGCG 693
 |||||
 OY 205 rGlnLysSerTyrProTyrLeuAlaGluAspGlnLysCysHisTyrLysAlaGlnAsnSe 225
 694 AGTGCACACATGTTCAAGATATGATGCTTCCCTTGGCAGTGAAGATGCTTAAAGAGA 753
 |||||
 Db 225 rGlyAlaThrAspLysGlyPheValAspIleGlnGluAlaAsnGlnLysAspLysAl 245
 754 AGCTGTGGCCCATTAAGAGCACTGTGCTGTGCGCTATAGATGCGAGCCACTATTCCTT 813
 |||||
 Db 245 dAlaValAlaThrValGlyProValSerIleAlaAlaAspAlaSerHisGlyIuThrGln 265
 814 CCTGATGAGAGTGGTGTCTACTATGAACTATCTGT---ACTTCAATGTGAATATGAG 870
 |||||
 OY 265 nLeuTyrSerAspGlyValIlyTyrSerAspProGlnCysSerSerGlnGlnLeuAsnHisn 285
 871 AGTATTACTGCTGCTTATGCTTAACCTT---AATGGGAAGAGTACCTGCTTGTGAAAA 927
 |||||
 Db 285 yValLeuValAlaGlyTyrGlyThrSerAspAspGlnGlnAspTyrTrpLeuValLysAs 305
 928 CAGCTGGCGCTCAACTTGTGTCACCAAGATATATGCGATGCGCAAGAAACATGCAAA 987
 |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 11:06:33 ; Search time 24 Seconds
(without alignments)
3891.862 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1997
Sequence: 1 gcatlataatgataagagca.....ataaatgtcatgacacat 1126

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame_n2p.model -DEV-xlp
-O/cgn2.1/USPTO_spool/US10010580/runat_27062003.104535.11721/app-query.fasta.1.1287
-DB-swissprot.40 -OEMT-fastlan -SUFFIX-n2p.rsp -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX-blosum62 -TRANS-human40.ccd
-LIST=45 -DOCALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10010580_@CGN_1_1_26_@runat_27062003.104535.11721 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEBUFFER -NEG_SCORES=0 -WAIT -DSPHLOCK=100 -LONCLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1568	78.5	331	1 CATS_HUMAN	P25774 homo sapien
2	1367.5	68.5	340	1 CATS_MOUSE	O70370 mus musculu
3	1334.5	66.8	330	1 CATS_RAT	Q03765 rattus norv
4	1089	54.5	329	1 CATS_BOVIN	P25326 bos taurus
5	944.5	47.3	329	1 CATK_RABIT	P43236 oryctolagus
6	943.5	47.2	329	1 CATK_MACRA	O77641 macaca fasc
7	939.5	47.0	329	1 CATK_HUMAN	P43235 homo sapien
8	937.5	46.9	329	1 CATK_RAT	O35186 rattus norv
9	933.5	46.7	330	1 CATK_PIG	O94181 sus scrofa
10	926.5	46.4	329	1 CATK_MOUSE	P55097 mus musculu
11	843	42.2	334	1 CSU2_HUMAN	O60911 homo sapien
12	835.5	41.8	334	1 CATL_MOUSE	P06797 mus musculu
13	823.5	41.2	334	1 CATL_RAT	P07154 rattus norv
14	812	40.7	334	1 CATL_BOVIN	P25975 bos taurus
15	806	40.4	333	1 CATL_HUMAN	P07711 homo sapien
16	804	40.3	334	1 CATL_PIG	O28944 sus scrofa
17	777	38.9	323	1 CYS2_HUMAN	P25782 homarus ame
18	767	38.4	333	1 CATJ_MOUSE	O94014 mus musculu

19	767	38.4	341	1 CATL_DROME	U95029 drosophila
20	759	38.0	321	1 CYS3_HOMAM	P25784 homarus ame
21	748	37.5	333	1 TEST_RAT	P15242 rattus norv
22	734	36.8	343	1 CATO_RAT	O94763 rattus norv
23	715.5	35.8	322	1 CYS1_ROMAM	P13277 homarus ame
24	689	34.5	344	1 CYS5_DICDI	P54640 dictyostell
25	685.5	34.3	218	1 CATL_CHICK	P09648 gallus gall
26	675	33.8	458	1 ORYA_ORYSA	P25776 oryza saliv
27	667	33.4	462	1 RD21_ARATH	P13297 arabidopsis
28	652	32.6	217	1 CATL_SHEEP	U10991 ovis aries
29	647	32.4	471	1 ORVB_ORYSA	P25777 oryza sativ
30	634	31.7	395	1 CATL_BROPA	O17473 brugia paha
31	617.5	30.9	376	1 CYS2_DICDI	P04989 dictyostell
32	616.5	30.9	328	1 CYS4_BRANA	P25521 brassica na
33	603	30.2	380	1 ACTN_ACTCH	P25785 actinidia c
34	592.5	29.7	362	1 CYS1_PHAVU	P25803 phascoglus v
35	587	29.4	371	1 CYS1_HORVU	P25249 hordeum vul
36	586	29.3	373	1 CYS2_HORVU	P25250 hordeum vul
37	577	28.9	362	1 CYS1_VITGMU	P12412 vitina mungo
38	573.5	28.7	236	1 CATU_RAT	O63088 rattus norv
39	573	28.7	221	1 GPIL_ZINOF	P82474 zingiber of
40	565	28.3	360	1 CYS2_MALZE	O10717 zea mays (m
41	564	28.2	360	1 CYS1_HEMSP	P43156 hemerocalli
42	563	28.2	346	1 CYS1_LYCDS	P20721 lycopersico
43	561.5	28.1	333	1 CATL_RAT	P00786 rattus norv
44	559.5	28.0	345	1 ANAN_ANNACO	P80884 ananas como
45	558	27.9	221	1 GP1_ZINOF	P82473 zingiber of

ALIGNMENTS

RESULT 1

ID	CATS_HUMAN	STANDARD;	PRT;	331 AA.
AC	P25774; Q9BUG3;			
DT	01-MAY-1992 (Ref. 22, Created)			
DT	15-JUN-2002 (Ref. 41, Last sequence update)			
DT	15-JUN-2002 (Ref. 41, Last annotation update)			
DE	Cathepsin S precursor (EC 3.4.22.27).			
GN	CTSS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Alveolar macrophage;			
RX	MEDLINE=92218373; PubMed=1373132;			
RA	Shi G.-P., Munger J.S., Meara J.F., Rich D.H., Chapman H.A.;			
RT	"Molecular cloning and expression of human alveolar macrophage			
RT	cathepsin S, an elastolytic cysteine protease.";			
RL	J. Biol. Chem. 267:7258-7262(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=94209337; PubMed=8157683;			
RA	Munger J.S., Webb A.G., Foster K.E., Knoll J.H.M., Lemere C.A.,			
RT	"Human cathepsin S: chromosomal localization, gene structure, and			
RT	tissue distribution.";			
RL	J. Biol. Chem. 269:11530-11536(1994).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92317106; PubMed=1377692;			
RA	Wiederanders B., Broemme D., Kirschke H., von Figura K., Schmidt H.,			
RT	Peters C.;			
RT	"Phylogenetic conservation of cysteine proteolases. Cloning and			
RT	expression of a cDNA coding for human cathepsin S.";			
RL	J. Biol. Chem. 267:13708-13713(1992).			
RN	[4]			
RP	REVISION TO 211.			
RC	TISSUE-Testis;			
RA	Wiederanders B., Broemme D., Kirschke H., von Figura K., Schmidt B.,			
RA	Peters C.;			

DB 321 GlycylalaserPhcProSerTyProGluIle 331

RESULT 2

OC CATS_MOUSE STANDARD; PRT: 340 AA.
 ID 070370; 054973;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Calpsein S precursor (EC 3.4.22.27).
 GN CATS OR CATS.
 OS Mus musculus (Mouse).
 OC Fukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
 OX NCBI_Taxid-10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/c, and 129/Sv; TISSUE-Brain;
 RA Doh-ura K.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Spleen;
 RA Romerskirch W.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 144-306 FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN-C57BL/6; TISSUE-Carilage;
 RA MEDLINE-99326135; PubMed-10395917;
 RA Socoderostron M., Salminen H., Glunoff V., Kirschke H., Aro H.,
 RA Vuorio R.;
 RT "Cathepsin expression during skeletal development.";
 RL Biochim. Biophys. Acta 1446:35-46(1999).
 RN [4]
 RP SEQUENCE OF 296-340 FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-Brain;
 RA MEDLINE-98184882; PubMed-9516475;
 RA Dandoy-Dron F., Guillo F., Benoudjema L., Deslys J.-P., Laemesas C.,
 RA Dornont D., Tovey M.G., Dron M.;
 RT "Gene expression in scrapie. Cloning of a new scrapie-responsive gene
 and the identification of increased levels of seven other mRNA
 transcripts.";
 RL J. Biol. Chem. 273:7691-7697(1998).
 CC -1- FUNCTION: THIOL PROTEASE. THE BOND-SPECIFICITY OF THIS PROTEINASE
 IS IN PART SIMILAR TO THE SPECIFICITIES OF CATHEPSIN L AND
 CATHEPSIN N.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin L, but with much less
 activity on Z-Phe-Arg-1-NHMe, and more activity on the Z-Val-Val-
 Arg-1-Xaa compound.
 CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION FOUND
 IN NON-SKELETAL TISSUES. RELATIVELY HIGH LEVELS FOUND IN SKELETAL
 TISSUES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: AF051732; AAC05781.1; -
 DR EMBL: AF051727; AAC05781.1; JOINED.
 DR EMBL: AF051728; AAC05781.1; JOINED.
 DR EMBL: AF051729; AAC05781.1; JOINED.
 DR EMBL: AF051726; AAC05781.1; JOINED.
 DR EMBL: AF051730; AAC05781.1; JOINED.
 DR EMBL: AF051731; AAC05781.1; JOINED.
 DR EMBL: AF038546; AAR84925.1; -
 DR EMBL: AJ002386; CAA05360.1; -

DR EMBL: Y18466; CAA77184.1; -
 DR EMBL: AJ223208; CAA11182.1; -
 DR HSSP: P25774; 1BXF.
 DR MEROPS: C01.034; -
 DR MGD: MGI:107341; Ctss.
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHprot_acstle.
 DR Pfam: PF00112; peptidase_C1; 1.
 DR PRINTS: PR00705; PAPAIN.
 DR PRODOM: PD000158; Peptidase_C1; 1.
 DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
 DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
 DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
 KW Hydrolase; Thiol protease; lysosome; signal.
 FT SIGNAL 1 17
 FT PROPEP 16 112
 FT CHAIN 113 340
 FT ACT_SITE 147 147
 FT ACT_SITE 287 287
 FT ACT_SITE 307 307
 FT DISULFID 134 233
 FT DISULFID 144 189
 FT DISULFID 178 222
 FT DISULFID 281 329
 FT DISULFID 120 120
 FT CARBOHYD 120 218
 FT VARIANT 218 218
 FT CONFLICT 1 29
 FT CONFLICT 34 34
 FT CONFLICT 97 97
 FT CONFLICT 106 106
 FT CONFLICT 146 146
 SV SEQUENCE 340 AA; 38438 MW; 06BE61126E2B0CDE CMC64;
 Alignment Scores:
 Pred. No.: 2.3e-114 Length: 340
 Score: 1367.50 Matches: 251
 Percent Similarity: 85.59% Conservative: 40
 Best Local Similarity: 73.82% Mismatches: 48
 Query Match: 68.48% Indels: 2
 DB: 1 Gaps: 1
 US-10-010-580-1 (1-1126) x CATS_MOUSE (1-340)
 QY 12 ATGGGAGCAGCAGGAGTTCATGAATGGTAGTGGAGCTCCCTGTCGTC 71
 DB 1 MetArgAlaProGlyHisAlaIleatgtrplepnetrmeuProLeuValCysSer 20
 QY 72 TATGAGTGGCAGCAAGTACATGAAGATCCCACTGGATCATCACTGCAAG 131
 DB 21 ValAlaMetGluGlnLeuGlnArgAspProThrLeuAspThrHisTrpAspLeuTrpLys 40
 QY 132 AAAAAGTACAGCAAAATC-AAGGAAGGATGAGGAGGAGTACAGGATCACTG 190
 DB 41 LysThrHisGluLysGluLysLysAspLysAsnGluGlnGluValArgArgLeuIleTrp 60
 QY 191 GAAAAAATATAAATTTGTGATGCTTCACAACTGGAACTCAATGGGAATGATTC 250
 DB 61 GluLysAsnLeuLysPheIleMetIleHisAsnLeuGluLysSerMetGlyMetHisThr 80
 QY 251 TATGATTCAGCATGAACCATCTGGAGACATCACTGCGAAGGATGATATCTTTG 310
 DB 81 TyrGlnValAlaIleMetAsnAspMetGlyAspMetLysAsnGluLysSerCysArgMet 100
 QY 311 GGTTCCTGAGAGTTCACAGCAATGGCAGAGAAATGTAACTTAATAGTTAAATCT 370
 DB 101 GlyAlaIleuValArgLieserArgInsSerProLysThrValThrPheArgSerTyrSerAsn 120
 QY 371 CAGAAATTCCTGATTTCTGTGAGTCTGAGAGTGAAGAGAAAGAGCTGTTTCTGA 430
 DB 121 ArgThrLeuProAspThrValAspTrpArgGluLysGluValCysValThrGluValLysTyr 140
 QY 431 CAGGATCTCTGTGTGCTTGTGTGGCTTTGAGCCCTCTGGGCCCTCGAAGACAC 490


```

DB 145 ValGAlaLeuGluGluGlnLeuLysLysLysThrGlyLysLeuLeuAsnLeuSerPro 164
UY 527 CAGACCTGCTAGCATTCCTCACTCAAGCAAAATATGCGCATAAAGCGCTCCAAATGGCGTTTC 586
DB 165 GlnmsnleuValAspCysValSerGlu-----AsnaspGlyCysGlyGlyGlyTyr 181
UY 587 ATGCAACCTGCTTTCCAGTATATATTGATACCAAGCGCATGATGAGAAAGCTTCCTAT 646
DB 182 MetThrAsnAlaPhoGlnTyrValGlnLysAsnArgGlyLeuSerGluAspAlaTyr 201
UY 647 CCGTACCAAGCGCATGCAATGCAAGCTGATGATGATGCAAAAAAGCGAGCTGCCCATGT 706
DB 202 ProTyrValGlnGlnGlnGlnSerCysMetTyrAsnProThrGlyLysAlaAlaLysCys 221
UY 707 TCAAGTATAGTCAACTTCCTTGGCAGTCAAGCATGCTTAAAGCAAGCTGTGGCAAT 766
DB 222 AtGcGlyTyrATGCluGlnTyrGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 241
UY 767 AAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
DB 242 ValGlyProValSerValAlaAlaAspAlaSerLeuThrSerPheGlnPheTyrSerLys 261
UY 827 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 883
DB 262 GlyValTyrTyrAspGlnSerCysAsnSerAspAsnLeuAsnAlaValLeuAlaVal 281
UY 884 GCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 943
DB 282 GcTyrTyrGlyTyrGlnLysGlyAsnLysHisTyrPheLeuLysAsnSerTyrGlyLys 301
UY 944 TTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1003
DB 102 TPrGlnAsnLysGlyTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 321
UY 1004 ATTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1027
DB 322 AsnLeuAlaSerPheProLysMet 329

RESULT 7
CATK_HUMAN
ID CATK_HUMAN STANBANK; PRK; 329 AA.
AC P43235;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin K precursor (Ref. 3, 4, 22, 38) (Cathepsin O) (Cathepsin X)
GN CTSK OR CTSO.
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE-Blood;
RX MEDLINE-95104457; PubMed-7805878;
RA Shi G.-P., Chapman H.A., Bhatia S.M., Delcuvu C., Reddy V.V.,
RA Weiss S.J.;
RT "Molecular cloning of human cathepsin O, a novel endoproteinase and
RT homologue of rabbit OC2.";
RL FEBS Lett. 457:129-134(1995).
RN 121
RP SEQUENCE FROM N.A.
RC TISSUE-Bone;
RX MEDLINE-95118380; PubMed-7818555;
RA Tsukada T., Bilbe G., Ishibashi O., Tezuka K.-I., Kamegawa M.,
RA Kokubo T.;
RT "Molecular cloning of human cDNA for cathepsin K: novel cysteine
RT proteinase predominantly expressed in bone.";
RL Biochem. Biophys. Res. Commun. 206:89-96(1995).
RN 131
RP SEQUENCE FROM N.A.
RC TISSUE-osteoclastoma;

```

```

RX MEDLINE-96172222; PubMed-8585423;
RA Li Y., Alexander M., Mucherplemign A.L., Yellick P., Chen W.,
RA Strashenko P.;
RT "Cloning and complete coding sequence of a novel human cathepsin
RT expressed in giant cells of osteoclastomas.";
RL J. Bone Miner. Res. 10:1197-1202(1995).
RN 141
RP SEQUENCE FROM N.A.
RC TISSUE-Spleen;
RX MEDLINE-96082523; PubMed-7576232;
RA Broemme D., Okamoto K.;
RT "Human cathepsin O2, a novel cysteine protease highly expressed in
RT osteoclastomas and ovary molecular cloning, sequencing and tissue
RT distribution.";
RL Biol. Chem. Hoppe-Seyler 376:379-384(1995).
RN 151
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE-97185906; PubMed-9033587;
RA McGrath M.E., Klaus J.L., Barnes M.G., Broemme D.;
RT "Crystal structure of human cathepsin K complexed with a potent
RT inhibitor.";
RL Nat. Struct. Biol. 4:105-109(1997).
RN 161
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE-98070737; PubMed-9405598;
RA Thompson S.K., Halbert S.M., Bossard M.J., Tomaszek T.A., Levy M.A.,
RA Zhao B., Smith W.W., Abdel-Meguid S.S., Janson C.A., D'Alessio K.J.,
RA McQueney M.S., Amegadzie B.Y., Hanning C.K., Jeszajans R.L.,
RA Briand J., Sarkar S.K., Huddleston M.J., James C.P., Carr S.A.,
RA Garner K.T., Shu A., Hays J.R., Bradbeer J., Zembryki D., Weber D.F.;
RT "Design of potent and selective human cathepsin K inhibitors that
RT span the active site.";
RL Proc Natl. Acad. Sci. U S A 94:14249-14254(1997).
RN 171
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF ZYMOGEN FORM.
RX MEDLINE-99110586; PubMed-9893980;
RA Lalonde J.M., Zhao B., Janson C.A., D'Alessio K.J., McQueney M.S.,
RA Orsini M.J., Debouck C.M., Smith W.W.;
RT "The crystal structure of human procathepsin K.";
RL Biochemistry 38:862-869(1999).
RN 181
RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
RX MEDLINE-99156066; PubMed-10048321.
RA Sivaraman J., Lalumiere M., Menard R., Cygler M.;
RT "Crystal structure of wild-type human procathepsin K.";
RL Protein Sci. 8:283-290(1999).
RN 191
RP VARIANT PYCNODYSTOSIS ARG-146.
RX MEDLINE-96355650; PubMed-8703060.
RA Gelb B.D., Shi G.-P., Chapman H.A., Desnick R.J.;
RT "Pycnodysostosis, a lysosomal disease caused by cathepsin K
RT deficiency.";
RL Science 273:1236-1238(1996).
RN 1101
RP VARIANT PYCNODYSTOSIS VAL-277.
RX MEDLINE-98198344; PubMed-9529353;
RA Gelb B.D., Willner J.P., Dunn T.M., Kardon N.H., Verloes A.,
RA Benzin J., Desnick R.J.;
RT "Paternal uniparental disomy for chromosome 1 revealed by molecular
RT analysis of a patient with pycnodysostosis.";
RL Am. J. Hum. Genet. 62:848-854(1998).
RN 111
RP FUNCTION: CLOSELY INVOLVED IN OSTEOCLASTIC BONE RESORPTION AND
RN 112
RP PARTICIPATE PARTIALLY IN THE DISORDER OF BONE REMODELING. DISPLAYS
RN 113
RP POTENT ENDOPEPTIDASE ACTIVITY AGAINST FIBRINOGEN AT ACID PH. MAY
RN 114
RP PLAY AN IMPORTANT ROLE IN EXTRACELLULAR MATRIX DEGRADATION.
RN 115
RP CATALYTIC ACTIVITY: Broad proteolytic activity. With small-
RN 116
RP molecule substrates and inhibitors, the major determinant of
RN 117
RP specificity is p2, which is preferably Leu, Met > Phe, and not
RN 118
RP Arg.
RN 119
RP SUBCELLULAR LOCATION: lysosomal.
RN 120
RP TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN OSTEOCLASTS (BONNS).
RN 121
RP DISEASE: DEFECTS IN CTSK ARE THE CAUSE OF PYCNODYSTOSIS, AN
RN 122
RP AUTOSOMAL RECESSIVE OSTEOCLASTOPATHY CHARACTERIZED BY

```


RX NCBI_TaxID=9606:
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=96222936; PubMed=9563472;
 RA Santamaria J., Velasco G., Cazotla M., Fueno A., Campo E.,
 Lopez-Olin C.;
 RT *cathepsin L2, a novel human cysteine proteinase produced by breast
 and colorectal carcinomas.*;
 RL Cancer Res. 58:1624-1630(1998).
 RN [2]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE=Corneal epithelium;
 RX MEDLINE=96394384; PubMed=9727401;
 RA Adachi W., Kawamoto S., Ohno I., Nishida K., Kinoshita S.,
 Matsubara K., Okubo K.;
 RT *Isolation and characterization of human cathepsin V: a major
 proteinase in corneal epithelium.*;
 RL Invest. Ophthalmol. Vis. Sci. 39:1789-1796(1998).
 RN [3]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC TISSUE=Kidney and Thymus;
 RX MEDLINE=9915210; PubMed=10029531;
 RA Broemme D., Li Z., Barnes M., Mehler E.;
 RT *Human cathepsin V functional expression, tissue distribution,
 electrostatic surface potential, enzymatic characterization, and
 chromosomal localization.*;
 RL Biochemistry 38:2377-2385(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99310345; PubMed=10382972;
 RA Itoh R., Kawamoto S., Adachi W., Kinoshita S., Okubo K.;
 RT *Genomic organization and chromosomal localization of the human
 cathepsin L2 gene.*;
 RL DNA Res. 6:137-140(1999).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX MEDLINE=20480081; PubMed=11027133;
 RA Somoza J.R., Zhan H., Bowman K.K., Yu L., Mortara K.D., Palmer J.T.,
 Clark J.M., McGrath M.E.;
 RT *Crystal structure of human cathepsin V.*;
 RL Biochemistry 39:12543-12551(2000).
 CC -1- FUNCTION: IS A CYSTEINE PROTEASE. MAY HAVE AN IMPORTANT ROLE IN
 CORNEAL PHYSIOLOGY.
 CC -1- CATALYTIC ACTIVITY: The recombinant enzyme hydrolyzes proteins
 (Serum albumin, collagen) and synthetic substrates (Z-Phe-Arg-
 NHmc > Z-Leu-Arg-NHmc; Lysosomal (Potential)).
 CC -1- SUBCELLULAR LOCATION: Lysosomal (Potential).
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN THE THYMUS AND
 TESTIS. ALSO EXPRESSED IN CORNEAL EPITHELIUM, AND TO A LESSER
 EXTENT IN CONDUCTIVAL EPITHELIUM AND SKIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
 CC
 CC This SWISS-Prot entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 or send an email to license@isb-sb.ch).
 CC -----
 CC EMBL Y14734, CAA75029.1; -;
 DR EMBL AB001928; BA25609.1; -;
 DR EMBL AF070448; AAC23598.1; -;
 DR EMBL AB019534; BAA34365.1; -;
 DR PDB 1FHO, 30-JUL-01.
 DR MEROPS: C01.009; -;
 DR GeneW: HGNC:2538; CTSL2.
 DR MIM: 603308; -;
 DR InterPro: IPR000668; peptidase_C1.
 DR InterPro: IPR000169; SHprot_acste.
 DR Pfam: PF00112; peptidase_C1; 1.
 DR PRINTS: PR00705; PAPA1N.

DR	ProDom: PD000158; Peptidase_C1; 1.	DR	ProSite: PS00139; THIOL_PROTEASE_CYS; 1.
DR	ProSite: PS00639; THIOL_PROTEASE_HIS; 1.	DR	ProSite: PS00640; THIOL_PROTEASE_ASN; 1.
KM	Hydrolase; Thiol protease; Glycoprotein; Lysosome; zymogen; Signal; 3D-structure.	KM	Hydrolase; Thiol protease; Glycoprotein; Lysosome; zymogen; Signal; 3D-structure.
FT SIGNAL	1	17	POTENTIAL.
FT PROPEP	18	113	ACTIVATION PEPTIDE.
FT CHAIN	114	334	CATHEPSIN L2.
FT ACT_SITE	138	138	
FT ACT_SITE	276	276	
FT ACT_SITE	301	301	
FT DISULFID	135	178	
FT DISULFID	169	211	
FT DISULFID	270	323	
FT CARBOHYD	221	221	
FT CARBOHYD	292	292	
FT CONFLICT	81	81	
SV SEQUENCE	334 AA; 37329 MW; CD2DE1AC5F242C8 CRC64;		
Alignment Scores:			
Pred. No.:	1,68e-67	Length:	334
Score:	843.00	Matches:	168
Percent Similarity:	65.36%	Conservative:	49
Best Local Similarity:	50.60%	Mismatches:	105
Query Match:	42.21%	Indels:	10
DR:	1	Gaps:	5
US-10-010-580-1 (1-1126) x CSL2_HUMAN (1-334)			
OY	54 CTGCTGCTCGTTGGTCTCTATCATGCTGCGACAGTACATGAA--GATCCCACTTGAT	110	111
DB	7 LENAIALAHAEYSLEUCLYLLEALASERIALVALPOLYSPRHEARGINASINLEUASR	26	26
OY	111 CATCACTGCAATCTCTGGAGAAACCTACACAAACATACAGAAACAAATGACCAAC	170	170
DB	27 Thrlystrpryglntlrprysalathrlthlsarqlq--Leulrglylaasnglug	46	46
OY	171 TAGCAGCGGCGTCATCTGGAGAAAAATCTAAATTTGTGATGCTTCACAACTTGGAAC	230	230
DB	46 LYTTPRALYAGALVALTTPRGLIULYASINMETYLEGLIULIUNHISASNGLYGI	66	66
OY	231 ACTCAATGGAAATGATTCATATGATCTAGGATATACATCTAGGATCTAGGATCTAGGAT	290	290
DB	66 YTSERGLNGLYLSHISGLYRHEPHEMELAMETASNALRPHOGLYASRHECTHISNG	86	86
OY	291 AAGAAGGATATCTTTGATGGTTCCTCCAGACTCCACAGCAATGGCAGCAAAATGTCA	350	350
DB	86 IUGLUPHEARGINMEIMELGICYSPHEARG--ASNDILNYPHEARGYSELGLYSV	105	105
OY	351 CTTAATAGSTCAACTTAATCAGAAATTCCTCATATTTCTGCACTGCAGACAGAAAGGCT	410	410
DB	105 ALPHEARGIULPRLPHELPHELEUASPLRHEUOLYSSERVALASPTPRGYLSGLYT	125	125
OY	411 GTGTACTGAAGTAATAACCGAGGTTCTTGCGGCTGTGGCTTTVAAGTATP	470	470
DB	125 YRVALTPHPRVALYASNGILYSGINCYSLYSCYSTRPRALRHESERIALRHRG	145	145
OY	471 GGGCCCTGGAACAGCAACTGAACCTAAAAACAGAAAGCTGGTGTCTGTAGTCCACAGA	530	530
DB	145 LVALLAENGUGLYGLIMELRPHARGYSTRGLYLSLEUVALSERLEUSERGLIHA	165	165
OY	531 ACTTGCTGATTTGCTCAACTCAAAAAATATGAGGATAAAGGATGTAATGCTGTTTATCA	590	590
DB	165 SNLEUVALASPCYSSETRARGPROLIN--GLYASINGLIYASINGLIYGLIYHEMELA	184	184
OY	591 CAACGCTTCAGATATTATGATACAAAGCAATATTAATTAATTTTATATCTT	650	650
DB	184 LARGALRPHOGLINTYVALYLSGLIUSNGLYGLIYLEUSPRERGLIUGLINSERTYPR	204	204
OY	651 ACAAAAGCATGATGAGAAATGCGACGATATGATCAAAAAAGCAGACCTGGCAATCTTCAA	710	710
DB	204 YRVALAVALAASPLIULIECYLSYSTRARGPRGLIUSINSETRVALAIAASINSPRHRG	224	224

1D CATL_HUMAN STANDARD; PRT; 333 AA.
AC P07711; Q960J0;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cathepsin L precursor (PC 3.4.22.15) (Major excreted protein) (MEP).
GN CTSL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
KP SOURCE: FROM N.A.
KA MEDLINE=88439405; PubMed=3421948;
RT Gal S., Gottesman M.M.;
RT Isolation and sequence of a cDNA for human pro-(cathepsin L).
RT Biochem. J. 253:303-306(1988).
KN 12)
KP SOURCE: FROM N.A.
KA MEDLINE=88417115; PubMed=2835398;
RT Joseph L.J., Chang L.C., Stamenkovich D., Sukhatme V.P.;
RT Complete nucleotide and deduced amino acid sequences of human and
RT murine procathepsin L. An abundant transcript induced by
RT transformation of fibroblasts.
RT J. Clin. Invest. 81:1621-1629(1988).
KN 13)
KP SEQUENCE FROM N.A.
KC TISSUE=Prostate;
KA Strausberg R.;
RT Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
KN 14)
KP SOURCE: FROM N.A.
KA MEDLINE=88137635; PubMed=3342869;
RT Riton A., Popovic T., Kotnik M., Machleidt W., Turk V.;
RT Amino acid sequences of the human kidney cathepsins H and L.
RT FEBS Lett. 228:341-345(1988).
KN 15)
KP SOURCE: FROM N.A.
KA MEDLINE=87174843; PubMed=3550705;
RT Joseph L.J., Lapid S., Sukhatme V.P.;
RT The major ras induced protein in NIH3T3 cells is cathepsin L.
RT Nucleic Acids Res. 15:3186-3186(1987).
KN 16)
KP SOURCE: FROM N.A.
KA MEDLINE=87127952; PubMed=3545185;
RT Mason R.W., Walker J.E., Northrop F.D.;
RT The N terminal amino acid sequences of the heavy and light chains of
RT human cathepsin L. Relationship to a cDNA clone for a major cysteine
RT proteinase from a mouse macrophage cell line.
RT Biochem. J. 240:373-377(1986).
KN 17)
KP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 18-333.
KA MEDLINE=97051808; PubMed=8896443;
KA Coulombe R., Grochulski P., Sivaraman J., Menard R., Mort J.S.,
KA Cygler M.;
RT Structure of human procathepsin L reveals the molecular basis of
RT inhibition by the prosegment.
RT EMBO J. 15:5492-5503(1996).
KN 18)
KP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 114-333.
KA MEDLINE=97286312; PubMed=9141479;
KA Fujishima A., Imai Y., Nomura T., Fujisawa Y., Yamamoto Y.,
KA Sugawara T.;
RT The crystal structure of human cathepsin L complexed with E-64.
RT FEBS Lett. 407:47-50(1997).
KN 19)
KP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 18-333.
KA Cygler M., Coulombe R.;
RT Submitted (Aug-1999) to the PDB data bank.
CC 1- FUNCTION: IMPORTANT FOR THE OVERALL DEGRADATION OF PROTEINS IN
CC LYSOSOMES.
CC 1- CATALYTIC ACTIVITY: Specificity close to that of papain. As
CC compared to cathepsin H, cathepsin L exhibits higher activity

CC towards protein substrates, but has little activity on Z-Arg-Arg-
CC NHMe, and no peptidyl-dipeptidase activity.
CC -1- SUBUNIT: DIMER OF A HEAVY AND A LIGHT CHAIN LINKED BY DISULFIDE
CC BONDS.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13451; CAA30981.1; -;
DR EMBL: M05256; CAA28877.1; -;
DR EMBL: M0496; AAA66974.1; -;
DR PIR: S01002; KKHU1.
DR PIR: A26069; A26069.
DR PDB: 1CJL; 12-AUG-97.
DR PDB: 1CS8; 23-AUG-99.
DR MEROPS: C01.032; -;
DR Genew: HGNC:2537; CTSL.
DR MIM: 116880; -;
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; SHProl_acctle.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS: PR00705; PAPA1N.
DR PRODOM: PD000158; Peptidase_C1; 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
KW Hydrolase; Thiol protease; Glycoprotein; Lysosome; Zymogen; Signal;
KW 3D-structure. 1
FT SIGNAL 1 17
FT PROPEP 18 113 POTENTIAL.
FT CHAIN 114 288 ACTIVATION PEPTIDE.
FT PROPEP 289 291 CATHEPSIN L HEAVY CHAIN.
FT CHAIN 292 333 CATHEPSIN L LIGHT CHAIN.
FT ACT_SITE 138 138
FT ACT_SITE 276 276
FT ACT_SITE 300 300
FT ACT_SITE 300 300
FT DISULFID 135 178
FT DISULFID 169 211
FT DISULFID 269 322 INTERCHAIN.
FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (IN REF. 3).
FT CONFLICT 56 56 M -> V (IN REF. 3).
FT CONFLICT 268 268 D -> N (IN REF. 4).
SO SEQUENCE 333 AA; 37564 MW; 8CD17D0EF859DB5 CRC64;

Alignment Scores:
Pred. NO.: 3.4e-64 length: 333
Score: 806.00 Matches: 163
Percent Similarity: 64.74% Conservative: 50
Best Local Similarity: 49.54% Mismatch: 96
Query Match: 40.36% Indels: 20
DB: 1 Gaps: 7

US-10-010-580-1 (1-1126) x CATL_HUMAN (1-333)
QY 75 GCATGGCACAAGTACATAAAGATCCGACTGTATCATATTAATTTTGAAGAA 134
DB 15 Alaseralathreuthrphasphisserleuc[ua]ac[ln]trhrlystrp[ys]aia 34
QY 135 ACCTACAGCAATACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 194
DB 35 Melhismnary--LeuTyrclymelasn[du]lun[gl]trp[al]g[al]a[al]t[tp]id 54
QY 195 AAAATCTAAAAATTGTCATCTTCACAAATCTGGAACCTCAATGGAATGATTCATATG 254
|||||:||||| :||| ||||| |||

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 11:18:34 ; Search time 104 Seconds
(without alignments)
4461.713 Million cell updates/sec

Title: US-10-010-580-1
Perfect score: 1997
Sequence: 1 gcatattatgtagggagca.....aataatggtcatgacat 1126

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame_n2p.model -DEV-xlp
-Q-/cqr2.1/USPTO_SPOOL/US10010580/runat_27062003.104536.11734/app_query.fasta.1.1287
-DB-SPTREMBL.21 -QFMT-fastan -SUFFIX-n2p.rspt -MINMATCH=0.1 -LOOPTCL=0
-LOOPTXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-Blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OFFMT-pio -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USP-CR10010580.acsn.1.1.138.runat.27062003.104536.11734 -NPU=6 -ICPU=3
-NO_MMP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LOMLAS
-DEV-TIMEOUT=120 -WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP ivirus:*
16: SP bacteriap:*
17: SP archcap:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1375.5	68.9	340	Q99M14	Q99M14 mus musculus

2	1016.5	50.9	331	13	Q90324	Q90324 cyprinus ca
3	865.5	43.3	338	13	Q90MC2	Q90MC2 orcohynehu
4	835.5	41.8	334	11	Q90DC0	Q90DC0 mus musculus
5	835.5	41.8	334	11	Q91U20	Q91U20 mus musculus
6	825	41.3	338	5	Q9Y0X2	Q9Y0X2 artemia san
7	814.5	40.8	333	6	Q9GKX8	Q9GKX8 ceratophyc
8	812	40.7	336	13	P79722	P79722 brachydanio
9	793.5	39.7	333	6	Q9G124	Q9G124 canis fami
10	793.5	39.7	334	13	Q90686	Q90686 gallus gall
11	792.5	39.7	316	5	Q8MSH3	Q8MSH3 rhodnius pr
12	781	39.1	313	5	Q27708	Q27708 nephrops no
13	775.5	38.8	334	11	Q91XK6	Q91XK6 mus musculus
14	772	38.7	333	11	Q9J181	Q9J181 mus musculus
15	770.5	38.6	338	5	Q46030	Q46030 stenohalid
16	768.5	38.5	337	5	Q45734	Q45734 caenorhabd
17	768	38.5	332	5	Q9NHB5	Q9NHB5 boophilus m
18	768	38.5	337	5	Q9NHB5	Q9NHB5 delia radlc
19	767.5	38.4	328	5	Q27759	Q27759 penaeus van
20	767	38.4	324	13	Q90499	Q90499 myxine glut
21	764	38.3	326	5	Q27760	Q27760 penaeus van
22	763.5	38.2	334	11	Q9J1A9	Q9J1A9 mus musculus
23	760.5	38.1	324	13	Q80FP6	Q80FP6 engraullis j
24	760.5	38.1	339	5	Q26636	Q26636 sarcophaga
25	760.5	38.1	344	5	Q26425	Q26425 bombyx mori
26	759	38.0	324	13	Q90498	Q90498 myxine glut
27	754.5	37.8	324	5	Q80FP7	Q80FP7 engraullis j
28	754	37.8	325	5	Q46152	Q46152 penaeus van
29	752.5	37.7	334	11	Q9ET52	Q9ET52 mus musculus
30	747	37.4	343	11	Q912F4	Q912F4 heterodera
31	747	37.4	374	5	Q18455	Q18455 heterodera
32	736.5	36.9	353	5	Q18456	Q18456 heterodera
33	730.5	36.6	333	11	Q912P3	Q912P3 mus musculus
34	730.5	36.6	333	11	Q912P3	Q912P3 mus musculus
35	728	36.5	333	5	Q9SVQ2	Q9SVQ2 frankliniell
36	727.5	36.4	331	5	Q46031	Q46031 sitophilus
37	726.5	36.4	333	11	Q9J196	Q9J196 mus musculus
38	724.5	36.3	331	5	Q46032	Q46032 sitophilus
39	724.5	36.3	331	11	Q912F2	Q912F2 frankliniell
40	720	36.1	333	5	Q9SVQ4	Q9SVQ4 mus musculus
41	716.5	35.9	329	11	Q9J184	Q9J184 mus musculus
42	709.5	35.5	311	5	Q8T529	Q8T529 lesciolela he
43	709.5	35.5	322	5	Q967D5	Q967D5 lesciolela he
44	709.5	35.5	326	5	Q24940	Q24940 fasciola he
45	707.5	35.4	311	5	Q9GRW5	Q9GRW5 fasciola he

ALIGNMENTS

RESULT 1
Q99M14 PREDIMINARY: PRT: 340 AA.
ID Q99M14
AC Q99M14
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cathepsin S.
GN CTSS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Eutelesomi:
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strusberg R.
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC002125; AA02125.1; -
DR HSSP; P25774; 18XF.
DR MEROPS; C01.034; -
DR MGD; MGI:107341; Ctss.
DR InterPro; IPR000668; Peptidase-C1.
DR InterPro; IPR000169; SHProl_acsite.
DR Pfam; PF00112; Peptidase-C1; 1.
DR PRINTS; PR00705; PAPAIN.

Prodom: p0000158; Peptidase_C1: 1.
 PROSITE: PS00640; THIOL_PROTEASE_ASN: 1.
 PROSITE: PS00139; THIOL_PROTEASE_CYS: 1.
 PROSITE: PS00639; THIOL_PROTEASE_HIS: 1.
 Hydrolase: Thiol protease.
 SQUENCE: 340 AA; 38456 MW; 076502611F319D87 CRC64:

Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 4.99e-120	340	252
Percent Similarity: 86.18%	Conservative: 41	
Best Local Similarity: 74.12%	Mismatches: 46	
Query Match: 68.88%	Indels: 2	
	Gaps: 1	

US 10 010-580-1 (1-1126) x 099M14 (1-340)

```

  12 ATGGCAGCAGCTGACGCTTCTTCATGCAATGGCTACTGGCCCTCTCCCTGCTGCTC 71
  1 MetArgAlaIleProGlyHisAlaAlaIleArgTrpLeuPheTrpMetProLeuValCysSer 20
  72 TATGACTGTAATATTAAGATATCTGATGATGATGATGATGATGATGATGATGATGAT 131
  21 ValAlaMetGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40
  132 AAAACCTACAGCAACAAATC-AAGCAAGCAATGAGAGTACGACGCGCTTCATCTGG 190
  41 LysThrHisGluLysGluTyrLysAspLysAsnGluGluValAlaArgAlaGluLeuIleTrp 60
  191 GAAAAAATCTAAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 250
  61 GluLysAsnLeuLysPheIleMetIleHisAsnLeuLysSerMetGlyMetHisThr 80
  251 TATGATCTAGCATGAGCAATCTGGAGACATGACTGATGATGATGATGATGATGATGATGAT 310
  81 TyrGluValGlyMetAsnAspMetGlyAspMetThrAsnGluGluLysLeuGluCysArgMet 100
  311 GCTTCCTGAGTACGCTTCCGACCAATGGCAGCAATGCTATAGCTAAATCTCAAT 370
  101 G1ValAlaLeuValIleProArgGlnSerProLysThrValThrPheArgSerTyrSerAsn 120
  371 CAGAAATTCGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 430
  121 ArgThrLeuProAspThrValAspTrpArgGluLysGluValThrGluValLysTyr 140
  431 CAGAGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490
  141 GlnGlySerCysGlyAlaCysTrpAlaPheSerAlaValGluAlaLeuGluGluGluGluGlu 160
  491 AAGCTAAACACAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
  161 LysLeuLysThrGluLysLeuLysLeuSerLeuSerAlaGluAlaAsnLeuValAspCysSerAsn 180
  551 - GAAAAATATGCAATTAAGCTGCAATGCGCTTTCATGCAACTGCTTTCAGTAT 607
  181 GluGluLysTyrGlyAsnLysGluCysGluGluGluGluGluGluGluGluGluGluGluGlu 200
  608 ATTATGATTAACAAATGATGATTAAGATTTGATTAACAAATGATTAACAAATGATTAACAAAT 667
  201 LysGluHisTyrLysAsnSerLysAsnArgAlaAlaThrCysSerAlaGlyTrpLeuGluPro 220
  668 AAGTCAGATATGATGCAAAAAACGAGCTGCGACATGTCGCAAGATATGATGCAACTGCTCC 727
  221 LysGluHisTyrLysAsnSerLysAsnArgAlaAlaThrCysSerAlaGlyTrpLeuGluPro 240
  728 TTTGCTCAGTACAGCTGCTTAAAAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
  241 PheGluLysPheLysAspAlaLeuLysGluAlaValAlaThrLysGluProValSerValGly 260
  788 ATATGATGAGAGCTATGCTTCTTCTGATGAGAGAGTGTGCTGATGATGATGATGATGATGATGAT 847
  261 LysAspAlaSerHisSerSerPhePhePheTyrLysSerGlyValLysAspAspProSer 280

```

```

  848 TGTACTGCAATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 907
  281 CysThrGlyAsnValAsnHisGlyValLeuValGlyTyrGlyThrLeuAspLysLys 300
  908 GACTACTGCTGCTGCAAAAAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 967
  301 AspTyrTrpLeuValLysAsnSerTrpCysLeuAsnPheGlyAspGluGlyTrpLeuArg 320
  968 ATGGCAAGAAACAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1027
  321 MetAlaArgAsnAsnLysAsnHisCysGlyLysAlaSerAspCysSerTrpProLysLeu 340

```

RESULT 2

PRELIMINARY: PRT: 331 AA.

```

  ID 090324
  AC 090324:
  DT 01-NOV-1996 (TREMblrel. 01, Created)
  DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
  DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
  DE Cysteine proteinase.
  OS Cyprinus carpio (Common carp).
  OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
  OC Cyprinidae; Cyprinus.
  NC NCBI_TaxId=7962;
  RN [1]
  RP SEQUENCE FROM N.A.
  RC TISSUE-OVARY.
  RA Tsai Y.-J., Huang F.-L.;
  RT "Molecular cloning of a cysteine protease from carp ovary."
  RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
  DR EMBL: L30111; AAA49207.1; -.
  DR HSSP: P25774; 1BXF.
  DR MEROPS: C01.034; -.
  DR InterPro: IPR000668; Peptidase_C1.
  DR InterPro: IPR000169; SHPTol_acidic.
  DR Pfam: PF00112; Peptidase_C1; 1.
  DR PRINTS: PR00705; PAPA1N.
  DR Prodom: P0000158; Peptidase_C1: 1.
  DR PROSITE: PS00640; THIOL_PROTEASE_ASN: 1.
  DR PROSITE: PS00139; THIOL_PROTEASE_CYS: 1.
  DR PROSITE: PS00639; THIOL_PROTEASE_HIS: 1.
  KW Hydrolase; Thiol protease.
  SQ SHQUENCE 331 AA; 36229 MW; 076DF34D32C1CF82 CRC64:

```

Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 1.89e-86 <td>331</td> <td>190</td>	331	190
Percent Similarity: 1016.50 <td>Conservative: 48</td> <td></td>	Conservative: 48	
Best Local Similarity: 74.84% <td>Mismatches: 76</td> <td></td>	Mismatches: 76	
Query Match: 59.75% <td>Indels: 5</td> <td></td>	Indels: 5	
	Gaps: 3	

US-10-010-580-1 (1-1126) x 090324 (1-331)

```

  66 TGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 125
  12 CysSerAlaAlaLeuAlaPhePheAsnTrpAsn-----LeuAspGlnHisTrpLeu 29
  126 TCGAAGAAACCTACAGCAACA-AATCAAGCAAGATGAGCAAGTACAGCAAGCTGCTC 184
  30 TrpLysLysThrHisAsnLysPheTyrSerSerLysAspGluGluGluGluGluGluGluGlu 49
  185 ATCTGGGAAAAAATCTAAATTTGTGATGCTTGCATCTGAGAACACTCAATGAGAAATG 244
  50 LeuTrpGluAlaArgAsnLeuLysLeuLysLeuHisAsnLeuGluLysAspLeuHisGlyLeu 69
  245 CATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 304
  70 HisSerTyrAspLeuGluLysLeuHisMetGlyLysMetThrThrGluGluGluGluGluGlu 89
  305 TTGATGAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 361

```

Dd		90	ThrlaualatThrlllgarValIProProcllyphelysarqcinlrrlaagiunphevalGly	109
Oy		362	AACCTCAATCACAAATTGGCCTGATTCGTGTGCAGTGGAGAGAAGGCGTGTTACTGAA	421
				:::
Dd		110	SerSerGlYAlalalaIleProaspserLeuaSPTRPARGluLysGlyTyIValSerSer	129
Oy		422	CTGAATAATACCAGCCCTTCCTGTGGTGCTTGGCGCTTCACAGCGCTGTGGGCCCTCGAA	481
Dd		130	ValllysasnIndlyalacysglsercystRPPAlapheserSerValGlyVAlaLeuglu	145
Oy		482	GCAACAACGTGAAGCTAAAAACAGGAAGAGTGTGTCTCTGAGTGCACAGAACTGGTAGAT	544
Dd		150	GlyGlnleumetylsThrTrnrclyLysLeuValaspleuSerProGlnAsnleuValasp	165
Oy		542	TGGTCACACTGAAAAAATATATGGATTAAMAGCGTCGAATGGCGGTTATATCAACACTGCTTC	601
Dd		170	CysSerSer---serTryrGlyasnTryrGlycysGlyGlyLeuMetSerAlaIalape	188
Oy		602	CAGTATATATATGATTAACAAGGCCATTGATTCAGAAAGCTTCTATCCCTACAAAACCATG	66
Dd		189	GlnTryrValllleaspaInglyLyIleasSergIserSetryrProTryrGluIyal	201
Oy		662	AATGAAAGTGCACATATGCATTCMAAAAACCGAGCGCCACATATGTCMAAGTATACTGAA	72
Dd		209	GlnclgylncysaTryrTryasnProserGlnleuAlalaAsncysThrlyrTryr	221
Oy		722	CTTCCTCTTGGCAGTGAAGATGCCTTAAAGAAAGCTGTGGCCANTAAAGAGCTGTCT	78
			:::	
Dd		229	ValArgInglyAspsrglugluAlaleuLySglnAlaValaIalasnlleclypProIleser	24
Oy		782	GGCGCTATAGATGTCGAGCGACTATTTCTTCCCTGACAGAAAGTGGTCTACTATGAA	84
Dd		249	ValAlaIleasPhalathHisproGlnPheIlelenTryargserGlyValTryfasnsp	26
Oy		842	CCATCCTTACTACAGATGTGAATCATCATGAGCATATATAGTGGTCATGGTAACTTAAT	90
Dd		269	ProserCysThrTrnAsnIleasnhisAlaValaIeulaValaGlyTryrGlyAlaIlea	28
Oy		902	GGGAAGAAGCTACTGGCTGTGAAAAACAGTGGGGCCTCAAGCTTGTGGTACCAAGCATAT	96
Dd		289	GlyGlnaspphetrPleuVallylasnsertPrplyrThrelpPheGlyAspGlyGlyTy	30
Oy		962	ATTGCGATGCGAAGAACAGTGAAGTACGTGGGANTGCTACTATTCCTCTT	1015
Dd		309	IleArqMeAlaIaArgasnGlnasnasnmelcysGlylleahIsthrProser	326
RESULT 3				
090WC2	ID	PRELIMINARY:	PRT;	338 AA.
090WC2				
AC	01-DEC-2001 (TEMBLrel_19, Created)			
DT	01-DEC-2001 (TEMBLrel_19, Last sequence update)			
DT	01-MAR-2002 (TEMBLrel_20, Last annotation update)			
US	Procethepsin L.			
DE	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Actinopterygii; Neoplethyrini; Teleostei; Euteleostei;			
OC	Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.			
OX	NCBI_TaxId=8022.			
RN	[1]			
RP	SOURCE: FROM N.A.			
KC	TISSUE-OVAR1;			
RA	Piat F., Kwon J.Y., Tyler C.R.;			
RT	*Molecular characterization of putative yolk processing enzymes and			
RT	their expression during oogenesis and embryogenesis in rainbow trout			
RT	Oncorhynchus mykiss.";			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF385668; AAK69706..1; ..			
DR	InterPro: IPR000668; Peptidase.C1.			
DR	InterPro: IPR000169; SHProl_acste.			
DR	Plam: PF00112; Peptidase.C1; 1.			
DR	ProDom: PD000158; Peptidase.C1; 1.			
DR	PROSITE: PS00640; THIOL_PROTASE_ASN; UNKNOWN_1.			

[illegible]


```

QY 576 ATGAGGCTTCATGACAACTGCTTTCCAGTATATATGATACAAAGCGCATGATGAG 635
    |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 179 snclglyleumectasptyrAlaIaphocIntyrAlaIaIaspsnclglyleuaspserg 199
QY 636 AAGCTTCCTATCCCTACAAAGCATGATGAGAAATGCGAGATATGACTCAAAAAAGCGAG 695
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 199 IugIserTyrProtyrGluAlaThrGluIuSerTyrAsnProGluTyrSerV 219
QY 656 CTGACCATGTCATGAATATAGTGAAGCTGCTTTGCGCACTGAAGTGCCTTAAAAAGAG 755
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 219 AlAlaAsnAspThrGlyPheValAspIlePro---LysGlnGluPheAlaLeuMetLysA 238
QY 756 CTGTGCTCAATTAAGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 815
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 238 IAlaAlaIaThrValGlyProIleSerValAlaIleAspAlaGlyHisGluSerPheMetP 258
QY 816 TGTAACAGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 872
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 258 hetyrLysGluGlyIleTyrPheGluProAspCysSerGluAspMetAspHisGly 278
QY 873 TATTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 920
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 278 AlLeuValValGlyTyrGlyPheGluSerThrGluSerAspAsnSerLysTyrTyrPheV 298
QY 921 TGAAGAAACAGTGGGCGCTCAACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 980
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 298 AlLysAsnSerTyrPheGlyGluIuTrpGlyMetGlyTyrIleLysMetAlaLysAsp 318
QY 991 GTGAAATCACTGTGATTCAGTATATCCCTTACCCGAGAAATC 1027
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 318 rGArgAsnHisCysGlyLeuAlaSerAlaAlaSerTyrProThrVal 333

RESULT 8
P79722 PRELIMINARY: PRT: 336 AA.
ID P79722:
AC 01-MAY-1997 (TREMBLREL. 03, Created)
DT 01-MAY-1997 (TREMBLREL. 03, Last sequence update)
DF 01-MAR-2002 (TREMBLREL. 20, Last annotation update)
DE Cathepsin L.
GN Hgql.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCHL_TaxID=7955;
OX
RN (1)
RP SEQUENCE FROM N.A.
RA Vogel A., Gerster T.;
RT "Expression of a zebrafish cathepsin L gene in anterior mesoderm
RT and hatching gland."
RL Dev. Genes Evol. 206:477-479(1997).
DR EMBL: Y08321; CAA69623.1; -.
DR HISSP: O60911; LFHO.
DR MEROPS: C01.032.
DR ZFIN: ZDB-GENE:980526-285; hgql.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; SHProt_acsite.
DR Pfam: PF00112; Peptidase_C1.
DR PRINTS: PR00705; PAPA1N.
DR ProDom: PD000158; Peptidase_C1.
DR PROSITE: PS00640; TH1OL_PROTHASE ASN. 1.
DR PROSITE: PS00139; TH1OL_PROTHASE_C1S. 1.
DR PROSITE: PS00639; TH1OL_PROTHASE_HIS. 1.
KW Hydrolase; Thiol protease.
SQ SEQUENCE 336 AA: 38029 MW: 5E795879F85F6FC2 CRC64:

```

Alignment Scores:

Pred. No.:	2,55e-67	Length:	336
Score:	812.00	Matches:	161
Percent Similarity:	65.93%	Conservative:	48
Best Local Similarity:	50.79%	Mismatches:	98

```

Query Match: 40.66% Indels: 10
DB: 13 Gaps: 5
us-10-010-580-1 (1-1126) x P79722 (1-336)
QY 96 GATCCCACTGTGATCATCATCTGAATCTCTGAGAAAAAGCTAGCAAGCAATCAAGG 155
    |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 21 AspIleGlnLeuAspAspHisTyrPheAsnSerTyrPheSerGlnHisGlySer--TyrH 40
QY 156 AAGAGAAATGAGCAAGTACGACGCGCTCTCATCTGCGAAAAAAATCTAAATTTTGATCC 215
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 40 IsgIAspValGluValGlyArgArgMetIleTrpGluGluAsnLeuArgLysIleGluG 60
QY 216 TTCACATCTGGAGCACTGAGATGAGATGATCATATATATGATGATGATGATGATGATG 275
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 276 GAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 332
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 80 LysAspMetThrAsnGluGluPheArgGlnAlaMetAsnGlyTyrHisAspProAsnG 100
QY 333 AATGCGACAGAAATGTCACTATAGTCAAACTCTAATGAGAAATGCGCATTCCTGCG 392
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 100 InThrSerGlnGlyProLeuPheMetGluProSerPhePheAlaIaProGlnGlnValA 120
QY 393 ACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 452
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 120 SPTPrArgGlnArgGlyTyrValThrProValLysAspGlnLysGlnCysGlySerCysT 140
QY 453 GGGCTTTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 512
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 140 rPhePheSerSerThrGlyAlaLeuGlnGlnLeuPheArgLysThrGlyLysLeuI 160
QY 513 TCTCTGAGTGCACAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 572
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 160 LeSerMetSerGlnGlnAsnLeuValAspCysSerArgProGln--GlyAsnGlnGlyC 179
QY 573 GGAATGCGGCTTTCATGACAACTGCTTCCAGTATATATGATGATGATGATGATGATGAT 632
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 179 YAsnGlyGlyLeuMetAspGlnAlaPheGlnTyrValLysGlnAsnLysGlyLeuAspS 199
QY 633 CAGAGCTTCTATCCCTACAAAGCATGATGGA--AACTGTATATATGCAATTAACAA 689
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 199 erGlnGlnSerTyrProTyrLeuAlaArgAspAspLeuProCysArgTyrAspProArg 219
QY 690 AGCGACCTGCCCATGTCATGCAAGTATGCACTTCCCTTGGCAGCAGCAATGCTTAA 749
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 219 heAsnValAlaLysIleThrGlyPheValAspIleProSerGlyAsnGlnLeuAlaLeu 239
QY 750 AAGAACTGTGCGCAATTAAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 809
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 239 eLAsnAlaValAlaAlaValAlaGlyProValSerValAlaIleAspAlaSerHisGln 259
QY 810 TCTTCTGTACAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 866
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 259 euGlnPheTyrGlnSerGlyIleTyrTyrGluArgAlaCysSerSerSerLeuLeuAspH 279
QY 867 ATGAGTATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 914
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 279 IAlaValLeuValAlaGlyTyrGlnGlnAlaAspValAlaGlyLysAsnArgTyrT 299
QY 915 GCGTGTGAAAAACAGCTGGGCGCTCACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 974
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 299 rIleValLysAsnSerTyrPheAspLysTyrGlyLysAspLysTyrIleTyrMetAlaI 319
QY 975 GAAGCAAGTGAATCACTGTGGATGATGATGATGATGATGATGATGATGATGATGATG 1021
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 319 YAspLysAsnAsnHisCysGlyValAlaThrLysAlaSerTyrPro 334

```

RESULT 9

ID	09GL24	PRELIMINARY:	PRT:	333 AA.
AC	09GL24:			

DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHprol_acsile.
 DR Pfam: PF00112; Peptidase_C1. 1.
 DR PRINTS: PR00705; PAPA1N.
 DR ProDom: PD000158; Peptidase_C1. 1.
 DR ProSITE: PS00640; THIOI_PROTEASE_ASN. 1.
 DR ProSITE: PS00139; THIOI_PROTEASE_CYS. 1.
 DR ProSITE: PS00639; THIOI_PROTEASE_HIS. 1.
 DR HydroLase: Thiol protease; Zymogen; Signal; lysosome; Glycoprotein.
 KW SIGNAL. 19
 FT PROPEP. 20
 FT CHAIN. 120
 FT ACT_SITE. 144
 FT ACT_SITE. 281
 FT ACT_SITE. 301
 FT DISULFID. 141
 FT DISULFID. 175
 FT DISULFID. 274
 FT CARBHYD. 108
 FT SEQUENCE. 334 AA: 37218 MW: DB3E23FBC21BF3B9 CRC64;

Alignment Scores:
 Prod. No.: 1,37e-65 Length: 334
 Score: 793.50 Matches: 162
 Percent Similarity: 62.80% Conservative: 49
 Best Local Similarity: 48.21% Mismatches: 116
 Query Match: 39.73% Indels: 9
 Gaps: 4
 DB:

US-10-010-580-1 (1-1126) x Q90686 (1-334)

QY 36 ATGAAATGGCTAGTGGCTGCTCCGTTGCTCCTATGAGAGGACAGATCATATAA 95
 DB 4 LeuHistrpeulaLeuLeuValLeuLeuLeuPrometAlaAlaAlaGlnLeuArPro 23
 QY 96 GATCCCACTGTGATCATCATCTGATCTGTGAGAAACCTACAGCAACAATCAAGG 155
 DB 24 GluProGluLeuAspAlaGlnIntrPaspLeuTrpLysArgThrIleGlnLysAlaValGln 43
 QY 156 AAGAGAAATGAGCAAGTACGACGGCGCT---CATTCGGGCAAAAAAATCTAAATTTGTGA 212
 DB 44 -ArgGlnGlyGlyArgAsnValProGluValAspLeuGlyGlnGluProGluValLHisAr 63
 QY 213 TGCTTCAACATCTGATCACTCAATGGAATGATTCATATGATCTAGCATGAAACATC 272
 DB 63 gCysProGlnArgGlyAla-ArgLeuGlyLysHisSerPheGlnLeuAlaMetAsnIlyTL 83
 QY 273 TGGGAGACATGATGCTGAGAAAGTATCTTGTAGTGGCTGCTGAGAGATTCCAGCC 332
 DB 83 euGlyAspMetLHisSerGlnValAlaValArgThrMetThrGlyLeuArGValArProArgS 103
 QY 333 AATGCCAGAGA-----AATCTCATATAGTTCAAACCTTAATCAGAAATTCCTCAT 386
 DB 103 eArArgProArGProAsnGlyThrLeuTrpValProAspTrpPseSerArgAlaArProAla 123
 QY 387 CTGTGAGCTGAGAGAGAGAGAGGCTGTGTTACTGAAAGTAAATCCAGGTTCTTGTG 446
 DB 123 laValAspTrpArgArgGlyGlyValThrProValLysAspGlnGlyIncysGlyS 143
 QY 447 CTGTGAGCTTTCAGCGCTGTGGGGCTGAGAGCAACAACCTAATAAAGAGAGAG 506
 DB 143 eArGStrpAlaPheSerSerValGlyAlaLeuGlnGlyGlnLeuLysArgTrpGlnIlyL 163
 QY 507 AGCTGCTGTCTGATGACAGCAAGCTGATGATGCTCAACTGAAAAATATGCGATA 566
 DB 163 yLeuLeuSerLeuSerProGlnAsnLeuValLtyCysValSer-----AsnAsn 180
 QY 567 AAGGCTCAATGCGGCTTCATGACAACCTGCTTTCATATATATATATATCAAGAGCA 626
 DB 180 snGlyCysGlyGlyGlyMetThrAsnAlaPheGlnLysValArgLeuAsnArgGlyI 200
 QY 627 TTGATTAGAAAGCTTCTATACCTACAAAGCCATGAGTGAAGAGTGCAGATATGACTCA 686

DB 200 leAspSerGluAspAlaTyrProTyrIleGlyGlnAspLysSerCysMetLysSerProT 220
 QY 687 AAAACGAGCTGCCACATGTTCAAAATATATCTCACTCCCTTGGCAGTCAAGATCCCT 746
 DB 220 hrGlyLysAlaAlaLysCysArgGlyLysArgGlnIleProGluAspAsnGlyLysAla 240
 QY 747 TAAAGAGCTGTGGCCCAATAAAGACCTGTGTGTGCTGATGATGATGATGATGATG 806
 DB 240 euLysArgAlaValAlaArgIleGlyProValSerValGlyLysAspAlaSerLeuProS 260
 QY 807 CTTCCTTCCTGACAGAGCTGTCTACTATGACCAATCGTACT---CACAAATGTA 863
 DB 260 eArPheGlnPheTyrSerArgGlyValLtyTrpAspThrGlyCysAsnProGluAsnLleA 280
 QY 864 ATCATGAGTATATAGCGTGTGCTATGCTAATGCTAATGCTAATGCTAATGCTAATGCT 923
 DB 280 snHisAlaValLeuAlaValAlaGlyGlyAlaGlnLysGlyThrLysHisLtrpIleLle 300
 QY 924 AAAACGCTGGGCGCTCAACTTGTGACCAAGATATATGATGATGATGATGATGATG 983
 DB 300 yAsnSerTrpGlyThrGluTrpGlyAsnLysGlyLtyValLeuLeuAlaArgAsnMetL 320

QY 984 GAATCAGCTGGGATTCCTAGTATACCTCTTACCCAGAAATC 1027
 DB 320 yGlnTrpCysGlyLleAlaAlaSerPheProLysMet 334

RESULT 11

Q8MSH3 PRELIMINARY; PRT; 316 AA.
 AC Q8MSH3: 2001-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Cathespine L-like cysteine protease.
 OS Rhodnius prolixus (Triatominae: Rhodnius).
 OC Eukaryota; Metazoa; Arthropoda; Insecta;
 OC Pterygota; Neoptera; Paraneoptera; Hemiptera; Euhemiptera;
 OC Heteroptera; Panheteoptera; Cimicomorpha; Reduviidae; Triatominae;
 OC Rhodnius.
 NC NCBI_TaxID=13249;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=INTESTINE;
 RA Lopez-Ordóñez T., Rodríguez M.H.;
 RT Characterization of a cDNA encoding a cathespine L-like protein of
 RL Rhodnius prolixus.";
 RL Insect Mol. Biol. 10:505-511(2001).
 DR EMBL: AF320565; AAL34984.1;
 DR InterPro: IPR000668; Peptidase_C1.
 DR InterPro: IPR000169; SHprol_acsile.
 DR Pfam: PF00112; Peptidase_C1. 1.
 DR PRINTS: PR00705; PAPA1N.
 DR ProDom: PD000158; Peptidase_C1. 1.
 DR ProSITE: PS00640; THIOI_PROTEASE_ASN; UNKNOWN_1.
 DR ProSITE: PS00139; THIOI_PROTEASE_CYS; UNKNOWN_1.
 DR ProSITE: PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
 KW protease.
 SO SEQUENCE. 316 AA: 35602 MW: B979F0FD1CE94722 CRC64;

Alignment Scores:
 Prod. No.: 1.68e-65 Length: 316
 Score: 792.50 Matches: 160
 Percent Similarity: 64.33% Conservative: 42
 Best Local Similarity: 50.96% Mismatches: 107
 Query Match: 39.68% Indels: 6
 Gaps: 3
 DB:

US-10-010-580-1 (1-1126) x Q8MSH3 (1-316)

QY 99 CCCACTCTGAT-----CATCACTGCAATCTGCGAAGAAACCTACAGCAACAA 149
 DB 4 ProSerPheAspLleAspProGlnGlnLtyrLeuAlaPheLysAlaMetHisGlyLysAsn 23


```

OY 605 TATATTATGAAACACGCGATTGATTCAGAAAGCTTCTTCTTACAAAGCATGAT 664
DB 173 TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT TTTT
OY 665 GGAAGATGCGATATGCTCAAAAAGAGGAGTCCCATGTTCAAGTACTACTGACT 724
DB 193 ArgserlysaTgphesapalalanserlelglalathrcystrhglyPhevalgluval 212
OY 725 CCCCTTGGCAGTACAGATCCCTTAAAGAACGCTGCGCCATTAAGACCTGCTGCTG 784
DB 213 Glnhst---ThrglgluAlalaleuhlsiglualvalserpslleglvalProilseval 231
OY 765 GCTATAGATGAGGACCATATTCTTCTCTGACAGAGTGGTGTCTACTATGACCA 844
DB 232 AalaleasPalaserHispheserPheclnpheryserserglYalTYTYGluLys 251
OY 845 TCCTGTACT---CACATGTCGATCATGAGTATTTAGTGTGGCTATGCTATGCTTAAT 901
DB 252 LyscysserProthrasnleuasphtlsglyalvalalaglyTYTYGlyThrGluSer 271
OY 902 GCGAAGACTACTGCTGCTGTGCAAAAAGAGTGGCGCCCTCAACTTGTGACCAAGATAT 961
DB 272 ThrgluasPtyrTrrpuevalylsasnsertrpglyserglYtrpglyaspalaglYtr 291
OY 962 ATTCGATGCGCAAGAAACAGTGGAAATCACTGTGGATTTGCTAGTATCCCTTACCCA 1021
DB 292 lilelymetSerArfgasnargaspasnscysglYlalealaserGluProserYtrPro 311
OY 1022 GAATC 1027
DB 312 ThrVal 313

RESULT 13
OY 091XK6 PRELIMINARY; PRT: 334 AA.
ID 091XK6:
AC 01-DEC-2001 (TREMblrel. 19, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DB Adult female placenta cDNA, KIKEN full-length enriched library,
DE clone:1600023H08, full insert sequence.
GN CTSJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN RP
RA STRAIN=C57BL/6J; TISSUE=PLACENTA;
RC Adachi J., Aizawa K., Akahira S., Akimura T., Aono H., Arai A.,
RA Arakawa T., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hangaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,
RA Komori H., Konda M., Koya S., Kurihara C., Matsuyama T., Okada T.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Okazaki Y., Okido T.,
RA Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D.,
RA Shibata K., Shibata Y., Shinagawa A., Shiraki T., Sogabe Y.,
RA Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,
RA Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SOURCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PLACENTA;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
RN [3]
RP SOURCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PLACENTA;
RX MEDLINE=99279253; PubMed=10349636;

```

```

RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.*;
RL Meth. Enzymol. 303:19-44(1999).
RN [4]
RP SOURCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PLACENTA;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Komori H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.*;
RL genome Res. 10:1617-1630(2000).
RN [5]
RP SOURCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PLACENTA;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Komori H., Akiyama J., Nishi K., Katsuna T., Tashiro H., Itoh A.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto P., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Matshiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.*;
RL genome Res. 10:1757-1771(2000).
DR EMBL; AK005526; BAB24099.1; -.
DR MEROFS; C01.038; -.
DR MGD; MGI:1349426; Ctsj.
DR InterPro; IPR000668; Peptidase_C1.
DR InterPro; IPR000169; SHpro_acsite.
DR Pfam; PF00112; Peptidase_C1_1.
DR ProDom; PD000158; Peptidase_C1_1.
DR PROSITE; PS00640; THIOLEPROTEASE_ASN; UNKNOWN_1.
DR PROSITE; PS00139; THIOLEPROTEASE_CTS; UNKNOWN_1.
DR PROSITE; PS00639; THIOLEPROTEASE_HIS; UNKNOWN_1.
DR SEQUENCE 334 AA; 37276 MW; A422B7679E76DD7C CRC64;

Alignment Scores:
Pred. No.: 6,61e-64 Length: 334
Score: 775.50 Matches: 159
Percent Similarity: 63.42% Conservative: 56
Best Local Similarity: 46.90% Mismatches: 99
Query Match: 38.83% Indels: 25
DB: Caps: 9

US-10-010-580-1 (1-1126) x 091XK6 (1-334)
OY 54 CTGCTTCGCTGCTGCTCTCTATGACGTG-----GCACAGTACATMAAGTCCCACT 104
DB 7 LeuLeuIleLeuGys---PheGlyValAlaSerGlyAlaGlnAlaHis---AspProLys 24
OY 105 CTGATCATCATCATGGAATCTCTGGAAGAAAACCTACAGCAAAACATCAAGCAAGCAATG 164
DB 25 LeuaspAlaGluTrrpLysAspTrpLysThrLysTyrAlaLysSerYtrSerProLys--G 44
OY 165 AGCAAGTACGACGCGGCTCATCTGCGGAAAAAATGTAAATTTGTGATCTTCACAAATC 224
DB 44 lncAlaAlaLeuTrrpTrrpAlaValAlaTrrpGluGluAsnMetArgMetlelGlyLysHisAsnL 64
OY 225 TGGAACACTCAATGGGAATGCATTCATATGATCTGAGATTAACATTTGGGAGCAATGTA 284
DB 64 yscLysasnserLeuGlyLysAsnAsnphethrMetLysMetAlaLysHisPheLyspGlnTr 84
OY 285 CTGTGACAGAGTATGATTTTGATGAGGCTCCCTGAGAGTCCAGCCCAATGCG----- 337
DB 84 hrSerGluLuphearGlySerlelIleAspAsnlelProlelProAlaLalMetThrAsp 104
OY 338 -----CAGAGAAATGTCACTTATAGCTCAAACTTAATACAGAAATGGCGCATTCG 389
DB 104 roHisAlaGlnAsnHisValSerlle-----GlyLeuProAspYtrL 118
OY 390 TGGAAGTGAAGAGAGAGGCGTGTACTGACGAAATACCGAGATTCTGTGGTGTCTT 449

```

```

Db 118 yAsprrprprgsluglglgyltYrValThProValArqasnlngllyscysglsySec 138
QY 450 GTTGGCCCTTTCACCCGCTGGGCGCCCTGGAGCAGCACTGACGCTAAAAACAGCAAAAC 509
Db 138 yATrAlaPhelAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 158
QY 510 TGGCTTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 569
Db 158 eutHrProLeuSerValGlnAsnLeuLeuAspCysSer---lystHrValGlnAsnlySG 177
QY 570 GCTTGCAATGCTGCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 629
Db 177 lYcysglInserglYThrAlaHlsGlnAlaPhelglutYrValLeuLysAsnlysglyLeuG 197
QY 630 ATTGACAGCTTCCTTCATCCCTACCAAGCCATGACATGACATGACATGACATGACATGAC 689
Db 197 lnaAlaGlnAlaHrThrYrProTylrGlglYlYsAspGlyProCysArqYrAlaYsSerGln 217
QY 690 AGCGAGCTTGCACATGCTTCAAGTATACTGAACCTCCCTTGGCAGTGAAGATGCCCTTA 749
Db 217 snAlaSerAlaAsnllrHrAspYrValAsnLeuPro---ProAsnGlnLeuYrLeuT 236
QY 750 AAGAGCTGTGCTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 809
Db 236 rValAlaAlaValAlaSerllleGlyProValSerAlaAlaAlaAlaAlaAlaAlaAlaAla 256
QY 810 TCTTCCTGTACAGAGCTGCTCTCTACTATGACCACTGCTGCTGCTGCTGCTGCTGCTGCT 866
Db 256 heArGpHeTylAsnGlylYlYlYrYrGlYrProAsnCysSerSerYrPhValAsnH 276
QY 867 ATGAGCTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 914
Db 276 lsaAlaAlaValAlaValAlaGlyYrGlYrGlySerGlYlYsAspAlaLysAspAlaAsnlyT 296
QY 915 GAGCTGTGAAAAACATGGAGCTCAAACTTGTGTGACCAAGATATATGCGAGGCA 974
Db 296 rPlleLlYsAsnSerTrpGlyGlnGlnTrpGlyMetAsnLylYrMetGlnlAlaAl 316
QY 975 GAAACAGCTGCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1027
Db 416 yAspHlsAsnAsnHlsCysGlylYlAlaAlaSerLeuAlaSerYrProAsnllle 333

RESULT 14
QY1181 PERLIMINARY: PRT: 333 AA.
AC 09J181:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE FPGS68 (Cathpsin 2 precursor).
CN CTS6 OR FPGS68 OR CTS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
OX NCBI-TaxID-10090;
RN 11
SEQUENCE FROM N.A.
KA Hemberger M.C., Himmelsbauer H., Ruchmann J., Zeitz C., Fundele R.H.;
KT "cDNA subtraction cloning reveals novel genes whose temporal and
KT spatial expression indicates association with trophoblast invasion.";
KL Dev. Biol. 0:0-0(2000).
RN 12
SEQUENCE FROM N.A.
KA Denning J.M., Konrad M., Reiman S., Werber I., Schwinde A.,
KT "Characterization of a cluster of cathpsin L-like cysteine peptidases
KT and related genes on mouse chromosome 13.";
KL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF250840; AAF81277.1; -
DR EMBL: AY014780; AAK00509.1; -
DR HSSP: P07711; ICOL.

```

```

DR MEMOPS: COL.031: -.
DR MGD: MGI:1860275; Cls8.
DR InterPro: IPR000668; Peptidase_C1.
DR InterPro: IPR000169; Shp1_acsite.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS: PR00705; PAPAin.
DR ProDom: PD000158; Peptidase_C1; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS_1.
KW Signal.
FT SIGNAL. 1 20 POTENTIAL.
FT CHAIN 114 333 CATHEPSIN 2.
SQ SEQUENCE 333 AA; 37803 MW; 57648FD20A9614E CRC64.

Alignment Scores:
Pred. No.: 1,4e-63 Length: 333
Score: 772.00 Matches: 157
Percent Similarity: 62.61% Conservative: 54
Best Local Similarity: 46.59% Mismatches: 106
Query Match: 38.66% Indels: 20
Caps: 7

US-10-010-580-1 (1-1126) x 09J181 (1-333)
QY 54 CTGCTTCGCTTGTCTCTCTATGACGTGCGCAAGTA-----CATAAAGATCCCTG 107
Db 6 LeuLeuAlaLeuLeuCysLeuGlyValAlaAlaGlnValThrGlnSerSerAspProSerLeu 25
QY 108 GATCATCATGCAATCTCTGCAAGCAAACTGACAGCAAACTGACAGCAAACTGACAGCA 167
Db 26 AspSerGlnTrpGlnGlnTrpGlnTrpGlnTrpGlnTrpGlnTrpGlnTrpGlnTrpGln 45
QY 168 AAGTAGCAAGAGAGCTGCTATTTGGAGAAAAATTTAAATTTGGATGCTTCAATCTG 227
Db 45 lYcylGlnLysArqAlaValAlaTrpGlnLysMetLysLeuValLysGlnHlsAsnllleG 65
QY 228 AAGCACTGAGGCAATGATCATGATCATGATCATGATCATGATCATGATCATGATCATGAT 287
Db 65 lYrYrAspGlnGlyLysLysAsnPhrThrMetLysAlaAlaAlaAlaAlaAlaAlaAlaAla 85
QY 288 CTGGAAGAGTCATATCTTTGATGCTTCCCTGAGCTTCCCAAGTCCCAAGTCCCAAGT 347
Db 85 lYcylGlnTrpYrArqYrMetLeuThrAspLleProValPro-----AsnP 100
QY 348 TCACCTTAGCTCAAACTGATCATGAAA-----TTCCTGATCTGTGACT 395
Db 100 heArGlyLysLysSerllleHlsGlnProllleAlaGlyTrpLeuProLysPhValAspT 120
QY 396 GAGAGAGAGAGGCTGTGTACTAGAGTAATATAGAGTCTTGTGTGTGTGTGTGTGTGTGT 455
Db 120 rParGlyArqGlyCysValThrProValLysAsnGlnGlyThrCysAsnSerCysTrp 140
QY 456 CTTCACCCCTGTGGGCGCCCTGGAGCAGCACTGAAAGCTAAAGACAGCAAGCTGCTGT 515
Db 140 lAphSerAlaAlaGlyAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 160
QY 516 CTGAGTGTGCAAGCACTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 575
Db 160 rLeuSerThrGlnAsnLeuValAlaPcySerArqLeuGln---GlyAsnPhelCysP 179
QY 576 ATGGCGGTTTCATGCAACTGCTTCCAGTAATATATGATTAATTAATTAATTAATTAAT 635
Db 179 heLysGlySerThrPhelLeuAlaLeuLysValTrpLysAsnArqGlyLeuGlnAlaG 199
QY 636 AAGCTTCTTATTCCTTCAAAAGCCATGATGCAAGTCAATGATGATGATGATGATGATGAT 695
Db 199 lYrSerThrYrProTylrGlyGlyThrAspGlyHlsCysArqYrHlsProGlnArqSer 219
QY 696 CTGCAATCTTCAAAATATAGTGAATCTCTTGGCAGTAAGTATGATGATGATGATGATGAT 755
Db 219 lAlaAlaGlylYrHrSerPhrSerPhrValSer---AsnSerGlyLysAspLeuMetArq 238
QY 756 CTGTGGCAATTAAGAGCTGTGTCTGTGCTATAGATGCGAGCACTATTTCTTCTCC 815

```

Db 238 laValaIahhrIleGIyPrOIleSerValGIyIleAspAlaIaGhNIaLysSerPheArgL 258

Oy 816 TGACAGAAAGTGGTGTCTACTATGACAACTCTGTACTACAGAAAT---GTCAATCATGGAG 872

Db 258 eauTyrrgIucIuGIyIleTyrryrrGIuPrOIlyScySerSerAsnIleIleAsnIleSerV 278

Oy 873 TATTAGGTGGTGCATGTGT-----AACCTTAATGGGAAGACACTACGGCTTG 920

Db 278 allenuValaIaGIyrrGIyrrGIyrrGIyrrGIyrrGIyrrGIyrrGIyrrGIyrrPheuI 298

Oy 921 TCAAAAACAGCTGGCCCTCAACTTTCGTGCACCAAGCATATATTCGGATCGCAAGAAACA 980

Db 298 IeLysasnSerNIIsGIyrrGIuGInrrPGLyMeLasnGIyrrMeLysLeuNIaIaArgIyLa 318

Oy 981 GTGCAATCATCTGTGGATTCGACAGTTCATCCCTCTTACCACGAATATC 1027

Db 318 rGAsnAsnNIIsScyScIyIleAlaSerTyrrAlaValTyrrPrrArgVal 333

RESULT 15

ID	046030	PREF. IMINARY;	PRT;	338 AA
----	--------	----------------	------	--------

DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
DE Cysteine proteinase
OS *Stilophilus zeamais* ('maize weevil')
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Phyllophaga; Curculionidae; Dryophthorinae; *Stilophilus*
OX NCBI_TaxID:70414

RP SEQUENCE FROM N.A.
 RX MEDLINE=97279038; PubMed=9133615;
 RA Matsuno T., Emori Y., Abe K., Arai S.;
 RT "Characterization of a gene family encoding cysteine proteinases of
 RT *Strophilus zeamais* (milne weevil), and analyses of the protein
 RT distribution in various tissues including alimentary tract and germ
 RT cells.";
 RL J. Biochem. 121:464-476(1997).
 DR FMHJ, DB2884; BAA24442.1; -.
 DR HSSP; 060911, 1FHO.
 DR MEROPS; c01_092; -.
 DR InterPro: IPR000169, SHProl_acste.
 DR Pfam; PF00112; Peptidase_C1; 1.
 DR PRINTS; PR00705; PAPA1N.
 DR ProDom; PD000158; Peptidase_C1; 1.
 DR PROSITE; PS00640; THIOL_PROTASE_ASN; 1.
 DR PROSITE; PS00139; THIOL_PROTASE_CYS; 1.
 DR PROSITE; PS00639; THIOL_PROTASE_HIS; 1.
 KW Hydroxase; thiol protease.
 SQ SEQUENCE 338 AA, 37570 MW, 104E1C34CF262937 CMC64;

Alignment Scores:

Pred. No.:	1.95e+63	Length:	338
Score:	770.50	Matches:	167
Percent Similarity:	63.25%	Conservative:	43
Best local Similarity:	50.30%	Mismatches:	110
Query Match:	38.58%	Indels:	13
DB:	5	Gaps:	5

US-10-010-580-1 (1-1126) x 046030 (1-338)

QY 56 GCTTCGTTTGTGCTCTTAAGCACTGGACACATACAAATAAAGCCCATCGCATATCA 115
|||::|||::|||::|||::
Db 8 AAlAlValValIleSerCysGlnAlaValSer-PhenylAsp---LeuValGIngluI 26

QY 116 CTGGAATCTCTGCACAACAAACCCTAACCAACCAATC-AAGGACAGACTATGCAAGTAGC 174
|||||::|||::|||::|||::
Db 26 nrrPseSerPheIysMetClmHiseryrSaasnTyAspSeriUnthrIucluaaygph 46

QY 175 AGCGGCTCTCATCTGGGAAAAAATCTAAATAATTGTGATGCTTCACANAITGGAGACACTC 234

Db	46	eatgmethylslePhehemelguasnaahstlslysalAlalyslnsiaslyleuphe	66
OY	235	AATGCCAATGCATTTCATATGATGCTAGGCATGAAACCATCGGAGACATGACTGGTGAACA	294
Db	66	rginglyphevallyspheulsleuilyleuasnylstyrAlaspmleuinhslisgl	86
OY	295	AGTGAATCTTGTCATG-----GGTCTCTGAGAGATTTCCTACCTA	333
Db	86	uphevalsetrThleuasnglypheasnystrhlyasnasnllleuleuylsglyseras	106
OY	334	ATGGCAGACAAATGCTACTTATAGCTCAACACTTAAATAGAAATGCTGTGAATCTGTGCA	392
Db	106	pleuasnaspsalvalaAlarphlelesterProalasnallylsleupProasprThvalas	126
OY	394	CTGCAGACAGACAGCGCTGTACTCAATGTAAGAAATACACGCTTTCTTGGCTGCTGTC	452
Db	126	CTTAAATAGAsrlysgfAlaValatrhrcllyvallyksasnglnlyhslscysgllyserCvstr	146

454 CGCTTTCAGCGCTGTGGGGCCCTGGAAGCACACTGAAGCTAAAAACAGGAAGCTGCT 513

Db 146 nserpheseraLatThrglSerleuglvglnHisphEARqLysThrglYlslcuaVa 166

514 GTCTTGAGTGCACAGAACTTGGTAGATTGCTCAACTGAAAAATATGGGATTAAGGCTG 573

Db 166 1SerLeuSerGluGlnAsnLeuValAspCysSer---GlyARQTYRGlyAsnAsnGlyCys 185

574 CAATGGCGTTTCATGACCACTGCTTCCAGTATATATTGATACCAACGGCATTCATTG 633

Db 185 sasngllycglyleumetaspsnalapheargtryrilelysaspsanglyglytleaspth 205

634 AGAAGCTTCCTATCCCTACAAAGCCATGAATGGAAGTGCAGATATGACTCAAAAAAGCG 693

205 RGIuYSSerTYrProTYrLeuAlaGIuAspGIuLysCysHisTYrIlysAlaGlnAsnSe 229

694 AGCTGCCACATGTTCAAAGTATACTGAACTTCCCTTGGCAGTGAAGATGCCCTAAAGA 75

Db 225 rGlyAlaThrAspLysGlyPheValAspIleGluGluAlaAsnGluAspAspLeuLysAl 241

754 AGCTGTGCCAATAAGGACCTGTCTGTGGCTATAGATGGGAGCCACTATTCTTTCTT 81

Db 245 aAlaVal¹Met²Thr³Val⁴Gly⁵Pro⁶Val⁷Ser⁸Ile⁹Ala¹⁰Ile¹¹Asp¹²Ala¹³Ser¹⁴His¹⁵Glu¹⁶Thr¹⁷Pro¹⁸Glu¹⁹ 26

814 CCTGTACAGAGTGGTGTCTACTATGACCATTCTGT---ACTCAGATGTGAATCATGG 877

Db 265 nlcuTYrSerAspGlyValTYrSerAspProGluCYsserSerGlnGluGluAspHisG1 28

871 AGTATTAGTGGTTGGCTATGTAACCTT---AATGGGAAGACTACTGGCTTGTGMAAA 92

Db 285 yvalLeuValIGlyTyrGlyThrSeraspaspGlyGlnaspTyrTrpLeuValLysas 30

928 CAGCTGGGCTCAACTTTGGTGACCAAGGATATATTCGGATGGCAAGAAACAGTGGAA 98

Db 305 userTrpgIyProserTrpgIyLeuAsngIyTYrIlleIysmetaIaArqAsngInaspAs 32

988 TCACCTGTGGATTGCTAGTTATCCCTCTTACCA 1021

Dd 325 nmetcysglyvalalaserGINalaserIyrpro 336

Search completed: June 27, 2003, 11:33:13

Job time : 115 secs

Downloaded from <http://ajph.org/> on November 10, 2015

Tue Jul 1 14:40:50 2003

us-10-010-580-1.n2p.ra1

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 27, 2003, 11:26:29; Search time 20.5 seconds
(without alignments)
3232.216 Million cull updates/sec

Title: US-10-010-580-1
Sequence: 1 gcatlataqlataggaagca.....ataaalgatgacat 1126

Scoring table:

BL0SUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters. 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=frame_plus.model -DEV=xdp
-O=/cgn2_1/USPTO/spool/US10010580/runat_27062003_104537_11774/app.query.fasta_1.1287
-DB=Issued_Patents_AA -QFMT=iaslan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCTL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=0 -NOH=ext -HEAPSIZE=500 -MTLEN=0 -MAXLEN=2000000000
-USER=US10010580.ecgn.1.1.28 @runat_27062003_104537_11774 -NCPU=6 -ICPU=3
-NO_MMAP -IARGCPU=0 -NGC_SCORE=0 -WAIT -DSEPHLOCK=100 -LONCLOG
-DEV_TIMECUT=120 -WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Backfilest.pep.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1569	78.6	331	1	US-08-208-007A-8
2	1569	78.6	331	4	US-09-701-685-2
3	1562	78.2	331	4	US-08-860-255A-2
4	1562	78.2	331	4	US-08-915-095A-8
5	1562	78.2	331	4	US-08-798-096-8
6	1562	78.2	331	4	US-08-798-095A-8
7	1560	78.1	331	1	US-08-330-121B-4
8	1560	78.1	331	5	PCT-US95-13820-4
9	944.5	47.3	329	1	US-08-208-007A-7
10	944.5	47.3	329	1	US-08-330-121B-3
11	944.5	47.3	329	4	US-08-915-095A-7
12	944.5	47.3	329	4	US-08-798-096-7

13	944.5	47.3	329	4	US-08-798-095A-7
14	944.5	47.3	329	5	PCT-US95-13820-3
15	939.5	47.0	329	1	US-08-330-121B-2
16	939.5	47.0	329	2	US-08-852-807-20
17	939.5	47.0	329	3	US-08-964-308-4
18	939.5	47.0	329	3	US-08-964-313-4
19	939.5	47.0	329	4	US-08-860-255A-1
20	939.5	47.0	329	4	US-09-069-138-4
21	939.5	47.0	329	4	US-08-915-095A-2
22	939.5	47.0	329	4	US-08-798-096-2
23	939.5	47.0	329	4	US-08-798-095A-2
24	939.5	47.0	329	5	PCT-US95-13820-2
25	937.5	46.9	329	2	US-08-806-959-2
26	937.5	46.9	329	1	US-08-208-007A-2
27	933.5	46.7	329	5	PCT-US94-04781-2
28	932.5	46.7	329	4	US-08-684-932A-36
29	930.5	46.6	329	3	US-08-964-308-15
30	930.5	46.6	329	3	US-08-964-313-15
31	930.5	46.6	329	4	US-09-069-138-15
32	929.5	46.5	329	3	US-08-964-308-14
33	929.5	46.5	329	3	US-08-964-313-14
34	929.5	46.5	329	4	US-09-069-138-14
35	835.5	41.8	334	3	US-09-515-039-64
36	835.5	41.8	334	4	US-09-515-039-64
37	832	41.7	334	3	US-08-883-526-1
38	806	40.4	333	1	US-08-208-007A-9
39	806	40.4	333	1	US-08-330-121B-5
40	806	40.4	333	3	US-08-883-526-3
41	806	40.4	333	3	US-09-120-365-65
42	806	40.4	333	4	US-09-515-039-65
43	806	40.4	333	4	US-08-860-255A-3
44	806	40.4	333	4	US-08-915-095A-9
45	806	40.4	333	4	US-08-798-096-9

ALIGNMENTS

RESULT 1
US-08-208-007A-8
Sequence 8, Application US/08208007A
Patent No. 5501969
GENERAL INFORMATION:
APPLICANT: HASTINGS, ET AL.
TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
NUMBER OF SEQUENCES: 14
CORRESPONDING ADDRESSES:
ADDRESSEE: CARILLA, BYRNE, BAIN, GILFILLAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/208,007A
FILING DATE: March 8, 1994
CLASSIFICATION: A35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
FILING DATE: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 8:

```

1 SEQUENCE CHARACTERISTICS:
2 LENGTH: 331 AMINO ACIDS
3 TYPE: AMINO ACID
4 STRANDEDNESS:
5 TOPOLOGY: LINEAR
6 MOLECULE TYPE: PROTEIN
7
8 US-08-208-007A B

```

Alignment Scores:

Pred. No.:	8.31e-167	Length:	331
Score:	1569.00	Matches:	2900
Percent Similarity:	93.35%	Conservative:	19
Best local Similarity:	87.61%	Mismatches:	22
Query Match:	78.57%	Indels:	1
DB:	1	Gaps:	0

US 10 010 580-1 (1-1126) X US-08-208-007A-8 (1-331)

[illegible]

Oy	875	TTTAGTGGTTGTCACCTTAATGGAAGAAGCACTGCTGTGTGAAAAAACAGCTGC	934
Db	281	LeuValValGlYglrGlAspLeuAsnGlnGlySgiutYrtPrLeuValLysAsnSerTrp	300
Oy	935	GCGCTAACACTTGCTACCAAGAGAAATAATCCGATGCCAAGAAACAGTCGAAATCATTC	994
Db	301	GlyHisAsnPhenIylcIucIucIylYrtIleArgMetAlaValArgAsnLysGlyAsnHisCys	320
Oy	995	GGGATTGCTAGTATTATGCTCTTACCAGCAAAATC	1027
Db	321	GlyIleValAserPheProserYrtPrProcIuuIle	331

RESULTS 2
HS-09-701-685-2

: Sequence 2, Application US/09701685
: Patent No. 6387629

; GENERAL INFORMATION:

APPLICANT: Schneider, Patrick
Yamamoto, Karon K

APPLICANT: French, Cynthia K.

APPLICANT: Reprogen, Inc.

;;	TITLE OF INVENTION: Use 81 calheps
;;	TITLE OF INVENTION: Endometriosis

FILE REFERENCE: 018002-001310US

CURRENT APPLICATION NUMBER: US/09/2001-03-26

PRIOR APPLICATION NUMBER: WO PCT/U

PRIOR FILING DATE: 1999-06-03

PRIOR FILING DATE: 1998-06-04

; NUMBER OF SEQ ID NOS: 4

```

; SOFTWARE: PatentIn Ver. 2.0
; SEO ID NO 2

```

; LENGTH: 331

TYPE: PRT
ORGANISM: Homo sapiens

US-09-701-685-2

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

Alignment scores:
Pred. No.: 8.17e-167

Score: 1569.00

Best local similarity: 87.61%

Query Match: 78.578

DB: 4

US-10-010-580-1 (1-1126) x US-09-701-

36 ATGAATGGCTAGTTGGCGTCCTCCGTTGTCCTCC

```

Db      1 MetLysArgLeuValCysValLeuLeuValCysSer
          |||||  |||||  ::|||  :::|||||

```

OY	96	GATCCCACTGTGCATTCACACTGGAATCTCTGGAGAAACCCTACACAACAATC - AAC	154
Dy	21	AspProThrLeuAspHisIstHrPhisLeuTrpLysLysThryTGilySGInTyrLys	40
OY	155	GAAGAGAATGAGGAGTAGCACGGCGTCTCATCTGGAAAAAATTMAATTGTGATG	214
Dy	41	GlutylasasnIudlvalavalartgargleouIetrgluylasasnleuylphvalmet	60
OY	215	CTTCACAATCTGGACAACACTCATGGAAATGCATTTATTTGATCTTACGATCAACCATCTG	274
Dy	61	LeuhIsasnleuSIuhIssetmelUymethIssetryraspleUglymetAsnhsyleu	80
OY	275	GGAACATGACTGGTGCAGAACAGTGAATCTTTGATGGTTCGCCAGAGTTCACACCAA	334
Dy	81	GlyAspmetLhisSerGluGluValmetSerLeumetSerLeuArgvalProSerGln	100
OY	335	TGGCACAAGAATGTCTACTTATAGTCAAACTTAAATCAAGAAATTGCTGTGATTCGTGGAC	394
Dy	101	TrpGlnIarGasnhIethThryLysSerasnProasnArglleleuProAspSerValAsp	120

QY	395	TTGACGACGAAGGCGCTGCTTACGTAAGGAAATACAGGGCTTGTGGCTTTGG	454
Db	121	TTpArGdluylSGlYcylsYalThrGluValylsYrGlnGlySerCylGlyAlaCysTrp	140
QY	455	GTTTTCAGCGTCTTGAAGATTTTGAAGATCAACTGAAGTTAAAGAAAGAGAGCTGGT	514
Db	141	AlaPheserAlaValAlcylAlaLeuGluAlaGlnLeuIleuIleuIleuIleuIleuVal	160
QY	515	TCGCTAGTGGCAATCAAACTGTGTAGATGGCTCAACTGAAATATGGCAATAAAGCGTGC	574
Db	161	SeLueSerAlaIaInsnLeuValAlaPscYserTrnGluIlyrGlyAsnIlySGlYcyls	180
QY	575	AATGGCGGTTTCATGTCAACTGCTCTTCCAGTATTATTTGATTAACAACGCGCATTTGATCA	634
Db	181	AsnGlyGlyPheMetThrThrAlaPheGlnIlyrIleIleAspAsnIlySGlYleAspSer	200
QY	635	GAAAGCTTCTATCTCTTCAAAATCATGAATGGAAGAGTACATATGATCTAAAAAACGGA	694
Db	201	AspAlaSerIlyrProIlyrIlysaIleMetAspLeuIlyrCysGlnIlyrAspSerIlyrIlyr	220
QY	695	GCGGCACATGTTCAGAAAGTATAGTCAACTGCTTGGCAGTGGAGAGATGCGCTTAAAGAA	754
Db	221	AlaIaladThCysSerIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr	240
QY	755	GCTGTGGCCATTAAGGACCTGTCTGTGGCTATAGATCGAGCCACTATTTCTTTC	814
Db	241	AlaValAlaAsnIlyrGlyProValSerValIlyrValAspAlaIlyrIlyrIlyrIlyrIlyr	260
QY	815	CTTACAGAAAGTGGTGTCTACTATGAACCATCTTACTACAGAAAGTGAATCATGAGATA	874
Db	261	LeuIlyrIlyrSerIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyrIlyr	280
QY	875	TTAGTGTGGCTATAGTAAATTTAAATGCGAAATTAATGATTTGATGAAAAAATATGTGG	934
Db	281	LeuValValAlcylGlyrGlyAspLeuAsnGlyIlyrGlnIlyrIlyrIlyrIlyrIlyrIlyr	300
QY	935	CGCCTCAACTTGGTGCACCAAGGATATTTCGATGCCAAGAAACAGTGGAAATCACACT	994
Db	301	GlyHisAsnGlyGlnGlnGlnGlnIlyrIlyrIleAlaPheAlaIlyrAsnIlySGlYAsnIlyCys	320
QY	995	GGGATTGCTAGTATACCTCTTACCCAGAAATC 1027	
Db	321	GlyIleAlaSerPheIlyrProGlnIlyr 331	
RESULT 3			
US-08-860-255A-2			
Sequence 2, Application US/08860255A			
Patent No. 6274336			
GENERAL INFORMATION:			
APPLICANT: Abdel-Meguid, Sherin			
APPLICANT: Desjarlais, Renee			
APPLICANT: Janson, Cheryl			
APPLICANT: Smith, Ward			
APPLICANT: Zhao, Baoquan			
TITLE OF INVENTION: Method of Inhibiting Cathepsin K			
FILE REFERENCE: P50574-X1			
CURRENT APPLICATION NUMBER: US/08/860,255A			
CURRENT FILING DATE: 1997-06-26			
PRIOR APPLICATION NUMBER: 60/008,108			
PRIOR FILING DATE: 1995-10-30			
PRIOR APPLICATION NUMBER: 60/007,473			
PRIOR FILING DATE: 1995-11-22			
PRIOR APPLICATION NUMBER: 60/008,992			
PRIOR FILING DATE: 1995-12-21			
PRIOR APPLICATION NUMBER: 60/013,748			
PRIOR FILING DATE: 1996-03-20			
PRIOR APPLICATION NUMBER: 60/013,764			
PRIOR FILING DATE: 1996-03-20			
PRIOR APPLICATION NUMBER: 60/013,747			
PRIOR FILING DATE: 1996-03-20			
PRIOR APPLICATION NUMBER: 60/017,455			
PRIOR FILING DATE: 1996-05-17			
PRIOR APPLICATION NUMBER: 60/017,892			

[illegible]

DB 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGluArgGluAspValLeuLysGlu 240
QY 755 GCTGGGCGCATTAAGACCGTCGTCGCGCTATGATCGAGCCCATCTCTTCTTC 814
DB 241 AlaValAlaAsnLysCylProValSerValGlyValAspAlaArgHisProSerPhe 260
QY 815 CTTCAGAAAGTGGTGTCTACTATGAACCATCTGTACTCAGAAATGTGAATCATGAGTA 874
DB 261 LeuTyrArgSerGlyValIleTyrTyrGluProSerCysThrGlnAsnValAsnHisCylVal 280
QY 875 TTACTGGTGGCTATGTGTAACCTTAATGGCAAGACACTACCGCTTGCGAAACACGCTGG 934
DB 281 LeuValValGlyTyrGlyValAspLeuAsnGlyLysGlyTyrTrpLeuValLysAsnSerTrp 300
QY 935 GCGTCACACTTGGTCAGACAGCATATATTCGATCGACAGAAACACGTGCAAAATCACTGT 994
DB 301 GlyHisAsnProGlyGluGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
QY 995 GGAATGCTAGTATCTCTCTTACCGCAAAATC 1027
DB 321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 4

US-08-915-095A-B

; Sequence 8, Application US/08915095A
; Patent No. 6384793
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTROCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PFI07D4
; CURRENT FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-915-095A-B

Alignment Scores:

Pred. No.: 4,95e-166 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Host Local Similarity: 87.01% Mismatches: 23
Query Match: 78.22% Indels: 1
DB: 4 Gaps: 0

US-10-010-580-1 (1-1126) x US-08-915-095A-B (1-331)

QY 46 ATGAATGCGTACCTTGGCGCTCCGCTGCGCTGCTATGCGACGTCAGCAAGTACATATAA 95
DB 1 MetLysArgLeuValCysValIleLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
QY 96 GATCCCACTTGTGATCTCATCTGATCTGTGCAAAATCAATACAGTAAATC-AAG 154
DB 21 AsprTrothricLeuAspHisHisIleThrLysIleThrLysIleThrLysGlnIleThrLys 40
QY 155 CAACAGATGACGACAGTACCAAGCGCGCTCATCTGCGAAATAATCTAAATTTTCATG 214
DB 41 GluLysAsnGluIleValAlaArgAlaLeuIleTrpLysAsnLeuLysPheValMet 60
QY 215 CTTCACATCTGCAATCTGCAATCTGCAATCTGCAATCTGCAATCTGCAATCTGCAATCTG 274
DB 61 LeuHisAsnLeuGluHisSerMetCylMetHisSerTrpAspLeuGluMetAsnHisLeu 80
QY 275 GGAACATGATGATGATGAAGATGATATTTTATGAGTACCTGAGAGATTCAGAGCA 334
DB 81 GlyAspMetThrSerGluGluValMetSerLeuThrSerSerLeuArgValIleProSerGln 100
QY 335 TGGACAGCAAAATGCTACTTATAGCTCAAACTTAATTCAGAAATTCCTGATCTCTGAC 394
DB 101 TrpGlnArgAsnIleThrTyrLysSerAspProAsnArgIleLeuProAspSerValAsp 120

QY 395 TGGACAGCAAAAGGCTGTGTATTA-TGAAGTGAANTATCAAGTCTTCTTCTGCTGG 454
DB 121 TrpArgGluLysGlyValIleThrGluValLysTyrGlnGlySerCysGlyAlaCysTrp 140
QY 455 GCTTCAGCGCTGTGAGGCTCTTCAAGTCAATCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 514
DB 141 AlaPheSerAlaValGlyValLeuGluAlaGlnLeuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTCTGAGTCGACACAGACTTGGTACATTCCTCAACTGAAAAATATGCGAATAAACCTGG 574
DB 161 ThrLeuSerAlaGlnAsnLeuValAspCysSerTrpGluLysTyrGlyAsnLysCylCys 180
QY 575 AATGGCGTTTCATGACAGCTGTTCCAGTATATATGATATGACAGCGTTATTC 634
DB 181 AsnGlyCylPheMetThrThrAlaPheGlnIleThrLeuAspAsnLysCylLysAspSer 200
QY 635 GAAGCTTCTATGCTTACAAAGCTATGAATGGAAGTGCATATATCAAAAAATCA 694
DB 201 AspAlaSerTyrProTyrLysAlaMetAlaSerGlnLysCysGlnTyrAspSerLysTrp 220
QY 695 GCTGGCAGCATGTCAGATATACGACTTGGCTTGGCAGTACGATGCTTAAAGAA 754
DB 221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu 240
QY 755 GCTGTGGCGCATTAAGACCGTCGTCGCGCTATGATCGAGCCCATCTCTTCTTC 814
DB 241 AlaValAlaAsnLysCylProValSerValGlyValAspAlaArgHisProSerPhe 260
QY 815 CTTCAGAAAGTGGTGTCTACTATGAACCATCTGTACTCAGAAATGTGAATCATGAGTA 874
DB 261 LeuTyrArgSerGlyValIleTyrTyrGluProSerCysThrGlnAsnValAsnHisCylVal 280
QY 875 TTACTGGTGGCTATGTGTAACCTTAATGGCAAGACACTACCGCTTGCGAAACACGCTGG 934
DB 281 LeuValValGlyTyrGlyValAspLeuAsnGlyLysGlyTyrTrpLeuValLysAsnSerTrp 300
QY 935 GCGCTCAACTTTGTGTACCAAGATATATTGAGATGCGCAAGAACAGTGAATCACTGT 994
DB 301 GlyHisAsnProGlyGluGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320

RESULT 5

US-08-798-096-B

; Sequence 8, Application US/08798096
; Patent No. 6387682
; GENERAL INFORMATION:
; APPLICANT: Hastings, et al.
; TITLE OF INVENTION: HUMAN OSTROCLAST-DERIVED CATHEPSIN
; FILE REFERENCE: PFI07D2
; CURRENT APPLICATION NUMBER: US/08/798,096
; CURRENT FILING DATE: 1997-02-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-798-096-B

Alignment Scores:
Pred. No.: 4,95e-166 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Host Local Similarity: 87.01% Mismatches: 23
Query Match: 78.22% Indels: 1
DB: 4 Gaps: 0

US-10-010-580-1 (1-1126) x US-08-798-096-B (1-331)

Applicant	Haslins, et al.	Title of Invention	HUMAN OSTEOCLAST-DERIVED CATHEPSIN
File Reference	PI107D3	Current Application Number	US/08/798,095A
Current Filing Date	1997-02-12	Number of Seq ID Nos	14
Software	Patentin Ver. 2.1	Seq ID No	8
Length	331	Type	PRT
Organism	Homo sapiens	Organism	Homo sapiens
US-08-798-095A-8			
Alignment Scores:			
Pred. No.:	4,95e-166	Length:	331
Score:	1562.00	Matches:	288
Percent Similarity:	93.05%	Conservative:	20
Best Local Similarity:	87.01%	Mismatches:	23
Query Match:	78.22%	Indels:	1
DB:	4	Gaps:	0
US-10-010-580-1 (1-1126) x US-08-798-095A-8 (1-331)			
QY	36 ATGAATGGCTAGTTGGGAGCTCTTCCTGCTGCTATCCAGTGGTCACAAAGTACATATAA	95	
Db	1 MetLysArgLeuValCysValLeuLeuValCysSerSerLeuValAlaGlnLeuHisLys	20	
QY	96 GATCCCACTCTGGATCATCACTGGAATCTCGAAGAAACCTACAGCAACAATC-AAG	154	
Db	21 AspProThrLeuAspHisHisThrIleuThrIleuThrIleuThrIleuThrIleuThrIleu	40	
QY	155 GAAGAGATGAGAGATGACACGGCGCTCATCTCGGAAAAAATCTTAAATTTGGTATG	214	
Db	41 GluLysAsnGluGluAlaValArgArgLeuIleThrGluLysAsnLeuLysPheValMet	60	
QY	215 CTTCACAATCTGGACACTCATGGAATGATATCATATGATCTTGACATGATCTG	274	
Db	61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerLeuValProSerGln	80	
QY	275 GGAACATGACTGCTGAACAGATGATCTTATGATGCTCCCTGAGAGTCCCAAGCCAA	334	
Db	81 GlyAspMetThrSerGluGluValMetSerLeuThrSerSerLeuArgValProSerGln	100	
QY	335 TGGCAGAGAAATGTCACTTATAGTCAAACTCAATGCAAAATTCGCTGATCTTGGAC	394	
Db	101 ThrIleAsnThrSerGluGluValMetSerLeuThrSerSerLeuArgValProSerGln	120	
QY	395 TGGAGAGAAAGGCGCTGTACTGGAAGTGAATACACAGGCTCTTGTGCTGTGTGG	454	
Db	121 ThrProGluLysGlyCysValThrGluValLysTyrGlnLysSerCysGlyAlaCysThr	140	
QY	455 GCTTTCACCGCTGTGGGGCCCTCGAAGCCAACTGAAGCTAAAGACAGAAAGCTGCTG	514	
Db	141 AlaPheSerAlaValAlaGluAlaLeuGluAlaGlnLeuLysLeuLysThrGlyLysLeuVal	160	
QY	515 TCTCTGAGTGCACAACTTGATGATGCTCAACTGAAAATATGCAATTAAGGCTG	574	
Db	161 ThrLeuSerAlaGlnAsnLeuValAspCysSerThrGluLysTyrGlyAsnLysGlyCys	180	
QY	575 AATGGCGGTTTCATACAGAGCTTCCAGATATATATGTAACAAAGGCATATGATTC	634	
Db	181 AsnGlyGlyPheMetThrThrAlaPheIleThrIleLeuAspAsnLysGlyIleAspSer	200	
QY	635 GAAGCTTCTCATCCCTACAAACCCATGATGCAAGAGTGCATATGCTCAAAAAACCG	694	
Db	201 AspAlaSerTyrProTyrLysAlaMetLysGlnLysGlnLysTyrAspSerLysTyrVal	220	
QY	695 GCTGCCACATGTTCAAAATATACATCACTCCCTTGTGAGTGAAGATGCTTAAAGAA	754	
Db	221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu	240	
QY	755 GCTGTGGCCATTAAGAGCTGTGTGTGGCTATAGATGGCAATCACTATCTTCTTC	814	


```

Db      118  Set1LcAspryArqylsYlscgilyvaIThrProvalLySaSngInclgIncgcgly 137
QY      446  GCTTTTGGCGGCTTTTAAATCTCTGAGAGCAACATAAATTAATTAAGAACAGA 505
Db      138  SerCystrPalAphSerSerValGlValLeugInglInLeuLYSLValSThgIly 157
QY      506  AACCTGCTGTCTTCAGTGCCACAGACATTCGTACAATGCCTCACACGAAAATATGGCAAT 565
Db      158  LysLeuLeuAlaSnleuSerProClnAsnLeuValAspCysValSerGIuaSnItty----- 175
QY      566  AAGAGCTGCATATGGCGTTTATGACACTGCTTTCGAGTATTATATGATTAACACGGC 625
Db      176  - ClysCysClYclYglYlYrMetThrAsnalApheClnTrYvalClnatGrasnarGily 194
QY      626  ATTGATTCATCAAAATCTTCTATATCTTA-AAAGCATGAATGGAAATGCCATATGACTCA 685
Db      195  ILAdSpSerClnAspAlaLayrProTyrgValIGlYglNAspRglUseCysMeLTyrAsnPro 214
QY      686  AAAAAGCGACCTGCCACATCTTCCAAAGTAATCTGACATCTCCCTTCGCACTGAAGATGCC 745
Db      215  ThrGlYslAlaLaILaLysCysArgIlyTYrTaGrGuIlleProGluGLYasnGILyula 234
QY      746  TTAAAGAAATCTTGCGTCATTAAGAGCTGTGTGTGGTATAGATGGCAGCCACTAT 805
Db      235  LeuLYsarGalAlaValAlaArgValGlYrProValSerValAlaAlaAspAlaSerLeuthr 254
QY      806  TTTTTTTCTTTGTAAAGATGATGCTTACTATATGTAAGCATCTGT- --ACTCAAGATGTG 862
Db      255  SerPheClnApheYrSerLysGilyalTYrTYrAspGluAnacySerSerAspAsnVal 274
QY      863  AATCATCGAGTATCTAGCTGTGGCGTAAGCTTAACCTTAATGGGAAAGACTAGCTGCTGTG 922
Db      275  AsnAlaValaValLeuAlaValaLgLYrGlyllegInLysGlyAsnLYSHstPrIdelle 294
QY      923  AAAAAGCTGGAGGCTTCAACTTGTGTACCAAGATATATATCGGATGGCAAGAAAGCT 982
Db      295  LysAsnSerTrpClyLysSerTrpClyLysAsnGilyrTLleuMeAlaIATrGAsnLYs 314
QY      983  GAAATATATTTGTGATTTGTCTATTAATCTTTCCTTAACCTCAAAATC 1027
Db      315  AsnAsnAlaCysGilylLeuAlaSnleuAlaSerPheProLYsMel 329

RESIDUE 15
US 08-330-121H 2
Sequence 2, Application US/08330121B
Patient No. 5746357
GENERAL INFORMATION:
APPLICANT: Okamoto, Dieter
APPLICANT: Okamoto, Kathleen
TITLE OF INVENTION: CATHEPSIN O2 PROTEASE
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehm, Honbach, Tesl, Albritton
ADDRESSER: 6 Herbert
STREET: Four Embarradero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patedit Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330,121H
FILING DATE: 27-Oct-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 48,304
REFERENCE/DOCKET NUMBER: A-60261/JUH/RMS
TELECOMMUNICATION INFORMATION:
```

[illegible]

```
OY 827 GGTCCTACTATGAACCATCTGT--ACTCAGAATGTAATCATGAGTATTAGTGGT 883
    |||||||
Db 262 G|YValIYrTYrAspG|uSerCysAsnSerAspAsnLeuAsnHisAlaValLeuAlaVal 281
    |||||||
OY 884 GCCTATGCTAACCTTATATGGCAAGACTACTGCGCTTGTAAGAAAACAGC|CGGGGCTCAAC 943
    |||||||
Db 282 G|YTYrG|Y|I|eG|nLySGLYAsnLySH|Strp|I|e|LySAsnSer|TrpG|Y|G|uAsn 301
    |||||||
OY 944 TTTCGTGACCAAGCATATATTCGATGGCAAGAAACAGTGAATCACA|CTGCGGATTGCT 1003
    ..|||
Db 302 TrpC|YAsnLySGLYTYr|I|e|LeuMetAlaTrpAsnLySAsnAsnAlaCysG|Y|I|GAla 321
    |||||||
OY 1004 AGTTATCCCTCTTACCCAGAAATC 1027
    :|:
Db 322 AsnLeuAlaSerPheProLySMet 329
    :|:
    |||
```

Search completed: June 27, 2003, 11:35:47
Job time : 29.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - protein search, using frame_plus.n2p model

Run on: June 27, 2003, 11:33:19 ; Search time 57.5 Seconds

(Without alignments)
4294.342 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1997

Sequence: 1 gcaatattagttatggtgaggaaataaagtgtcatgacat 1126

Scoring table: BLASTSUM62

Xgapop 10.0, Ygapext 0.5

Ygapop 10.0, Ygapext 0.5

Delop 6.0, Delext 7.0

Searched: 424699 seqs, 10646833 residues

Total number of hits satisfying chosen parameters: 849398

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+.n2p.model -DEV=xip

-O=/cgn2_1/USPRO.spool/US10010580/runat_27062003.104539.11849/app_query.fasta_1.1287

-DB=Published.Applications_AA -OPMT=fastan -SUFFIX=n2p.rapb -MIMATCH=0.1

-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62

-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR_SCORE=pcr -THR_MAX=100

-THR_MIN=0 -ALIGN=15 -MODE=local -OUTFMT=fig -NCRM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=200000000 -USER=US10010580 -CGN_1_1_17.#runat_27062003.104539.11849

-NCPU=6 -ICPU=3 -NC_MMAP -LRS:EUCERY -NEG_SCORES=0 -WAIT -DSBLCK=100

-LONGIOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FCGAPOP=6 -FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Published.Applications_AA:

1: /cgn2_6/p/odata/1/pubpaa/US08_NEW_PUB.pep.*

2: /cgn2_6/p/odata/1/pubpaa/FCF_NEW_PUB.pep.*

3: /cgn2_6/p/odata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/p/odata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/p/odata/1/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/p/odata/1/pubpaa/US07_PUBCOMB.pep.*

7: /cgn2_6/p/odata/1/pubpaa/PCFUS_PUBCOMB.pep.*

8: /cgn2_6/p/odata/1/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/p/odata/1/pubpaa/US09_NEW_PUB.pep.*

10: /cgn2_6/p/odata/1/pubpaa/US09_PUBCOMB.pep.*

11: /cgn2_6/p/odata/1/pubpaa/US10_NEW_PUB.pep.*

12: /cgn2_6/p/odata/1/pubpaa/US10_PUBCOMB.pep.*

13: /cgn2_6/p/odata/1/pubpaa/US60_NEW_PUB.pep.*

14: /cgn2_6/p/odata/1/pubpaa/US60_PUBCOMB.pep.*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	1775	88.9	331	9	US-10-010-580-2
2	1569	78.6	331	9	US-10-009-275-2
3	1565	78.4	331	9	US-09-990-064-2
4	1562	78.2	331	10	US-09-953-956-8

5	1562	78.2	331	12	US-10-114-464-8	Sequence 8, Appli
6	1545.5	77.4	330	9	US-10-010-577-2	Sequence 2, Appli
7	944.5	47.3	329	10	US-09-953-956-7	Sequence 7, Appli
8	944.5	47.3	329	12	US-10-114-464-7	Sequence 7, Appli
9	939.5	47.0	329	10	US-09-953-956-2	Sequence 2, Appli
10	939.5	47.0	329	12	US-10-114-464-2	Sequence 2, Appli
11	939.5	47.0	361	10	US-09-925-300-1693	Sequence 1693, Ap
12	843	42.2	334	9	US-10-028-072-12	Sequence 12, Appl
13	843	42.2	334	9	US-10-121-049-12	Sequence 12, Appl
14	843	42.2	334	9	US-10-123-904-12	Sequence 12, Appl
15	843	42.2	334	9	US-10-140-470-12	Sequence 12, Appl
16	843	42.2	334	9	US-10-175-746-12	Sequence 12, Appl
17	843	42.2	334	9	US-10-176-918-12	Sequence 12, Appl
18	843	42.2	334	9	US-10-176-921-12	Sequence 12, Appl
19	843	42.2	334	9	US-10-137-865-12	Sequence 12, Appl
20	843	42.2	334	9	US-10-140-474-12	Sequence 12, Appl
21	843	42.2	334	9	US-10-142-431-12	Sequence 12, Appl
22	843	42.2	334	9	US-10-143-114-12	Sequence 12, Appl
23	843	42.2	334	9	US-10-140-002-12	Sequence 12, Appl
24	843	42.2	334	9	US-10-142-419-12	Sequence 12, Appl
25	843	42.2	334	9	US-10-123-262-12	Sequence 12, Appl
26	843	42.2	334	9	US-10-142-423-12	Sequence 12, Appl
27	843	42.2	334	9	US-10-121-050-12	Sequence 12, Appl
28	843	42.2	334	9	US-10-141-755-12	Sequence 12, Appl
29	843	42.2	334	9	US-10-143-031-12	Sequence 12, Appl
30	843	42.2	334	9	US-10-123-108-12	Sequence 12, Appl
31	843	42.2	334	9	US-10-123-236-12	Sequence 12, Appl
32	843	42.2	334	9	US-10-123-261-12	Sequence 12, Appl
33	843	42.2	334	9	US-10-140-921-12	Sequence 12, Appl
34	843	42.2	334	9	US-10-140-928-12	Sequence 12, Appl
35	843	42.2	334	9	US-10-121-045-12	Sequence 12, Appl
36	843	42.2	334	9	US-10-123-292-12	Sequence 12, Appl
37	843	42.2	334	9	US-10-123-903-12	Sequence 12, Appl
38	843	42.2	334	9	US-10-124-819-12	Sequence 12, Appl
39	843	42.2	334	9	US-10-124-825-12	Sequence 12, Appl
40	843	42.2	334	9	US-10-140-925-12	Sequence 12, Appl
41	843	42.2	334	9	US-10-160-498-12	Sequence 12, Appl
42	843	42.2	334	9	US-10-121-041-12	Sequence 12, Appl
43	843	42.2	334	9	US-10-121-041-12	Sequence 12, Appl
44	843	42.2	334	9	US-10-121-047-12	Sequence 12, Appl
45	843	42.2	334	9	US-10-123-215-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-10-010-580-2
; Sequence 2, Application US/10010580
; Publication No. US20030104972A1
; GENERAL INFORMATION:
; APPLICANT: Unimond, robin 1
; APPLICANT: Baker, sherry
; APPLICANT: Karlsson, lars
; TITLE OF INVENTION: polynucleotide and polypeptide sequences of canine cathepsin s
; FILE REFERENCE: ORT-1458
; CURRENT APPLICATION NUMBER: US/10/010,580
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: canine
US-10-010-580-2

Alignment Scores:
Pred. No.: 8.09e-180
Score: 1775.00
Percent Similarity: 99.70%
Best Local Similarity: 99.70%
Query Match: 88.88%
DB: 9
Length: 331
Matches: 330
Conservative: 0
Mismatch: 1
Gaps: 0

US-10-010-580-1 (1-1126) x US-10-010-580-2 (1-331)

```

UY 16 ATGAATGGCTAGTGGCTGCTCCGTTGTCCTCATGAGTGGCAAGTACATATAA 95
Db 1 MetLysTrpLeuValGlyLeuLeuProLeuLysSerLysAlaValAlaGlnValHisLys 20
UY 96 GATCCACATCTGGATATATCTGGAATCTGCAAGAAACCTACAGCAACAATC-AAG 154
Db 21 AspProThrLeuAspHisHisThrAsnLeuTrpLysLysThrLysSerLysGlnLysLys 40
UY 155 GAAACGAAATGGGAAATGAGTAACTCTTCTCATTTGGTAAAAAATCTTAAATTTGTATG 214
Db 41 GlnGlnAsnGlnGlnGlnAlaArgArgLeuLeuTrpGlnLysAsnLeuLysPheValMet 60
UY 215 CTTCACAAATCTGCAACATCATATGGGATCATATATATATAGGATGAAACATCTG 274
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerLysAspLeuGlnMetAsnHisLeu 80
UY 275 GGACACATGACTGGTGAAGAGTCAATCTTGAATGGCTTCCCTGACAGTCCAGCCAA 334
Db 81 GlyAspMetThrGlnGlnGlnValLysSerLeuMetGlySerLeuArgValProSerGln 100
UY 335 TGCCAGCAAAATGCTCATATATAGTCAAACTTAACTCAAAATGGCTGATTCGTGGAC 394
Db 101 TrpGlnArgAsnValThrTrpArgSerAsnSerAsnGlnLysLeuProAspSerValAsp 120
UY 395 TGAATGAGAGAGAGAGTGTGTTACTGAAATGCAAGAGGCTTCTGTTGCTTGTGG 454
Db 121 TrpArgGlnLysGlnLysValThrGlnValLysLysGlnLysSerGlnLysLysLys 140
UY 455 GCTTTCAGCGCTGTCGGCGCCCTGCAAGCAACTGCAAGCTAAAAACAGCAAGCTGCTG 514
Db 141 AlaPheSerAlaValGlyAlaLeuGlnAlaGlnLeuLysLeuLysTrpGlnLysLeuVal 160
UY 515 TCTCTGAGTGCACAGAAATTTGATGATGCTCAACTGAAATATGGAATTAAGAGCTGC 574
Db 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysLysLysGlnLys 180
UY 575 AATGGCGGTTTCATGACAACTGCTTCCAGTATATTTATTAATAACAAGGCAATGATTC 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnLysLysLysLysLysLysLysLysLys 200
UY 634 GATCCACATCTGCAACATCTGCAAGAGTCAATGCAAGAGTCAATGCAAGAGTCAATG 694
Db 201 GlnLysSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 220
UY 695 GCTGATCATATATTAATATATATGAACTTCCCTTGGCAGTGAAGATGCTTAAAGAA 754
Db 221 AlaAlaThrCysSerLysLysLysLysLysLysLysLysLysLysLysLysLys 240
UY 755 GCTTTCAGCGCTGTCGGCGCCCTGCAAGCAACTGCAAGCTAAAAACAGCAAGCTG 814
Db 241 AlaValAlaAsnLysGlnLysLysLysLysLysLysLysLysLysLysLysLys 260
UY 815 CTTCACAGAAATGCTCATATAGTCAAACTGCAAGAGTCAATGCAAGAGTCAATG 874
Db 261 LeuTrpArgAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 280
UY 875 TTATGAGTGTGCTATGATATTAATGCAAGAGTCAATGCTTGGTGAAGAGCTG 934
Db 281 LeuValValGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 300
UY 935 GCTTTCAGCGCTGTCGGCGCCCTGCAAGCAACTGCAAGCTAAAAACAGCAAGCTG 994
Db 301 GlnLysSerAlaGlnAsnLeuValAspCysSerThrGlnLysLysLysLysLys 320
UY 995 GCGATTCCTAGTATATCCCTCTTACCCAGCAATC 1027
Db 321 GlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 331

```

RESULT 2
US-10-099-275-2
Sequence 2, Application US/10099275

```

; Publication No. US20020187499A1
; GENERAL INFORMATION:
; APPLICANT: Schneider, Patrick
; APPLICANT: Yamamoto, Karen K.
; APPLICANT: French, Cynthia K.
; APPLICANT: Reprogen, Inc.
; TITLE OF INVENTION: Use of Catchpoin S in the Diagnosis and Treatment of
; FILE REFERENCE: 018002-001310US
; CURRENT APPLICATION NUMBER: US/10/099,275
; CURRENT FILING DATE: 2002-04-13
; PRIOR APPLICATION NUMBER: US/09/701,685
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/098,017
; PRIOR FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-099-275-2

```

```

Alignment Scores:
Pred. No.: 6,45e-158 Length: 331
Score: 1569.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.57% Indels: 1
Gaps: 0

```

US-10-010-580-1 (1-1126) x US-10-099-275-2 (1-331)

```

UY 36 ATGAATGGCTAGTGGCTGCTCCGTTGTCCTCATGAGTGGCAAGTACATATAA 95
Db 1 MetLysTrpLeuValGlyLeuLeuProLeuLysSerLysAlaValAlaGlnValHisLys 20
UY 96 GATCCACATCTGGAATATCTGGAATCTGCAAGAAACCTACAGCAACAATC-AAG 154
Db 21 AspProThrLeuAspHisHisThrAsnLeuTrpLysLysThrLysSerLysGlnLysLys 40
UY 155 GATCCACATCTGGAATATCTGGAATCTGCAAGAAACCTACAGCAACAATC-AAG 214
Db 41 GlnLysSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 60
UY 215 CTTCACAGAAATGCTCATATAGTCAAACTGCAAGAGTCAATGCAAGAGTCAATG 274
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerLysAspLeuGlnMetAsnHisLeu 80
UY 275 GGACACATGACTGGTGAAGAGTCAATCTTGAATGGCTTCCCTGACAGTCCAGCCAA 334
Db 81 GlyAspMetThrSerGlnGlnValLysSerLeuMetSerLeuArgValProSerGln 100
UY 335 TGCCAGCAAAATGCTCATATAGTCAAACTGCAAGAGTCAATGCAAGAGTCAATG 394
Db 101 TrpGlnArgAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 120
UY 395 TGAAGAGAGAGAGTGTGTTACTGAAATGCAAGAGTCTTGGTGAAGAGCTG 454
Db 121 TrpArgGlnLysGlnLysValThrGlnValLysLysLysLysLysLysLysLys 140
UY 455 GCTTTCAGCGCTGTCGGCGCCCTGCAAGCAACTGCAAGCTAAAAACAGCAAGCTG 514
Db 141 AlaPheSerAlaValGlyAlaLeuGlnAlaGlnLeuLysLeuLysTrpGlnLysLeuVal 160
UY 515 TCTCTGAGTGCACAGAAATTTGATGATGCTCAACTGAAATATGGAATTAAGAGCTGC 574
Db 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysLysLysLysLys 180
UY 575 AATGGCGGTTTCATGACAACTGCTTCCAGTATATTTATGATTAACAAGGCAATGATTC 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnLysLysLysLysLysLysLysLysLys 200

```

[illegible]

```

: RESU.LT3
: US-09-990-064-2
: Sequence 2, Application US/09990064
: Patent No. US20020164765A1
: GENERAL INFORMATION:
: APPLICANT: Parvinder K. Punia
: TITLE OF INVENTION: NOVEL COMPOUNDS
: FILE REFERENCE: GP-30242
: CURRENT APPLICATION NUMBER: US/09/990,064
: CURRENT FILING DATE: 2001-11-21
: PRIOR APPLICATION NUMBER: GH0028462.0
: PRIOR FILING DATE: 2000-11-22
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO: 2
: LENGTH: 331
: TYPE: prt
: ORGANISM: Mammalia
: US-09-990-064-2

```

Alignment Scores:	
Pred. No.	1 72e-157
Score:	1565.00
Percent Similarity:	91.84%
Best Local Similarity:	87.01%
Query Match:	78.37%
DB:	9
Gaps:	0

QY ATCAAAATGGCTAGCTGGAGCTCTTGGCTCTCTCAAGCGGCGCAAGTACATAAA 95
 Db 1 MellystrpLeuValItrpAlaLeuLeuValCysSerSerThrValAlaGlnLeuHisArg 20
 QY GATCCACCTCTGGATATCATCTGGATATCTCTGAGCAAAACCTTACACGCAAACTAC 154
 Db 21 AsproThrLeuAspHisHisIstrHisLeuTrpLysAlaIatGlyGlnItyLys 40
 QY GAGAGAGATATAGAGAACTATAGAGAGCTTATCTATCTAGAGAAAAAATTAATTTGTATG 214
 Db 41 GlutysAsnGlnCuaIaIaIatGATgLeuIleItrpGlnLysAsnLeuLysPheValThr 60
 QY 155 GAGAGAGATATAGAGAACTATAGAGAGCTTATCTATCTAGAGAAAAAATTAATTTGTATG 214
 Db 215 CTTCACAACTGGCAACACTCAATGGAGAAAGCACTTCAATATGATCTGGGCTGAACCAITGG 274
 Db 61 LeuHisAsnLeuGlnHisSerLysGlyMetHisIleSerTyrAspValGlyMetAsnHisLeu 80

QY	275	GGGACATACATCTTTTAAAGAAATGATATTTTATATATTTCTTGAAGTTTAAACCA	334
Db	81	AlaAspMetLthSerGluValValaSerLeuMetSerSerLeuAlaGlyLeuProHisGln	100
QY	335	TGCGAAGAAATGTCCTACTATTAAGTCAAATCTATATATATAATATGATATTTCTGGAC	394
Db	101	TrpProAlaGlnValThrTrpLysLeuAsnProAsnGlnLysLeuProAspSerValAsp	120
QY	395	TGCAGACACAGAGCGCTGTACTGTAAGTCAAAATACACAGCGTCTCTTGCTGCTTTCG	454
Db	121	TrpArgGluArgGlyCysValaThrGluValaLysTrpGlnLysSerCysGlyAlaAcySTP	140
QY	455	GGTTTAGCGCGTGGGGCGCTGGAGACACAACTGAAGCTAAAAACAGGAAAGCTGATG	514
Db	141	AlaPheSerAlaValaGlyAlaLeuGluAlaGlnLeuLysLeuLysThrGlyAsnLeuVal	160
QY	515	TCCTGATGTCACACAGAACTTGCTAGATTGCTCAACTATAAATATGACAAATACATGC	574
Db	161	SerLeuSerAlaGluAsnLeuValaAspCysSerThrTrpLysTrpGlyAsnLysGlyCys	180
QY	575	AATGGCGGTTTCATGCAACCTGCTTTCCAGTATATATTATTCATACCAACGGCATTTGATTC	634
Db	181	AsnGlyGlyPheMetThrGluAlaPheGlnTrpIleLeuAspAsnAsnGlyIleAspSer	200
QY	635	GAAGCTTCCTATCCCTACCAAAAGCTTCAGATGGAAGAGTGCATATATGATCAAAAAATCTA	694
Db	201	GluAlaSerTrpProTrpLysAlaMetAspGlnLysCysHisTrpAspSerPheHisArg	220
QY	695	GATGCAATATGTTCAAAAT	754
Db	221	AlaAlaThrCysSerLysTrpThrGluLeuPheProHegLysSerGluGluAlaLeuLysGlu	240
QY	755	GGCTGGCCAAATTAAGGACCTGTGCTGTGGCTATAGATGGCAGCCACTATCTTTCTTC	814
Db	241	AlaValAlaAsnLysGlyProValSerValAlaIleAspAlaSerHisSerSerPhePhe	260
QY	815	CTGACACGAAGTGGTCTACTACTATGACCATCTGTACTACGANTGGAAATCAATCAGAGTA	874
Db	261	LeuArgArgSerGlyValaLysTrpGluTrpSerCysThrGluAsnValaAsnHisGlyVal	280
QY	875	TTAGTGGTTGGCTATGTAACCTTAATGGGAAGAAGTACTGGCTGTGTGAAAAACAGCTGG	934
Db	281	LeuAlaValaGlyTrpGlyAsnLeuLysGlyLysAspTrpTrpLeuValLysAsnSerTrp	300
QY	935	GGCCTCAACTTTGGTACCAAGAGATATATCGATGGCAAAACAGTGCAAATCACTGT	994
Db	301	GlyIleHisPheGlyGlnGlnLysTrpIleArgMetAlaArgAsnSerLysAsnHisCys	320
QY	995	GGGATCTGTATGATCCCTCTTACCCAGAAATC 1027	
Db	321	GlyIleAlaAsnTrpProSerTrpProGluIle 331	

```

: RESULT 4
: US-09-953-956-B
: Sequence 8, Application US/09953956
: Patent No. US20020072107A1
: GENERAL INFORMATION:
: APPLICANT: Hastings, et al.
: TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
: FILE REFERENCE: P1107D2D1
: CURRENT APPLICATION NUMBER: US/09/953,956
: CURRENT FILING DATE: 2001-09-18
: PRIOR FILING DATE: 1998-12-23
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 8
: LENGTH: 331
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-953-956-B

```

ORGANISM: Homo sapiens
US-09-953-956-8

TYPE: PRT
ORGANISM: Homo sapiens
US-09-953-956-7

Alignment Scores:

Prod. No.:	1,61c-91	Length:	329
Score:	944.50	Matches:	186
Percent Similarity:	69.55%	Conservative:	47
Host Local Similarity:	55.52%	Mismatches:	79
Query Match:	47.30%	Indels:	24
DB:	10	Gaps:	5

US-10-010-580-1 (1-1126) x US-09-953-956-7 (1-329)

54 CTCTCCGCTGCTGCTCTATGACGTGACGACCAATACATTAAGATCCACTGATCAT 113
|||||
8 LeuLeuProValValSerPheAla-----LeuHisProGluGluIleLeuAspThr 24

114 CACTGCAATTTTGAAGAAATTTAGAGAAATCAAGAAAGAAAT-GAGGAAGTA 172
|||||
25 GlnTrpGluLeuTrpLysLysThrLysSerLysGlnLysAsnSerLysValAspGluIle 44

173 GCACGGCGCTGATCGGCAAAAAAATCTAAATTTGCTGATGCTGACATCTGGAAAC 232
|||||
45 SerArgGlnLeuLeuTrpGluLysAsnLeuLysHisLysSerLysHisAsnLeuGluAla 64

243 TCATGCAATGATCATATGATGATGATGATGATGATGATGATGATGATGATGAT 292
|||||
65 SerLeuGluValHisThrLysLysLysLysLysLysLysLysLysLysLysLysLys 84

293 GAATGATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 352
|||||
85 GluValValGlnLysMetThrGlyLeuLysValProProSer-----CCTGAT 98

353 TATAGTCAAACTCTAATCAAAATG-----CCTGAT 385
|||||
99 ArgSerHisSerAsnAspThrLeuTrpLysLeuProAspTrpGluLysArgThrProAsp 117

186 TTCTGCACTGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 445
|||||
118 SerLysAspTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 137

446 GCTTGGGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 505
|||||
138 SerGlySerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 157

506 AATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 565
|||||
158 LysLeuLeuAsnLeuSerProGlnAsnLeuValAspLysValSerGluAsnTrp----- 175

566 AAGCGCTGCAATGCGCTTTCATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 625
|||||
176 GlySerGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 194

626 ATTGATCAAG 685
|||||
195 LeuAspSerGlnAspAlaTrpProLysValGlyGlnAspLysSerGlyMetLysPro 214

686 AAAAGCGACGCTGACATCTTCAAACTATGATGATGATGATGATGATGATGATGAT 745
|||||
215 ThrGlyLysAlaAlaLysGlySerGlyLysLysLysLysLysLysLysLysLysLys 234

746 TTAAGCAAGCTGCGGCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805
|||||
243 LeuLysArgAlaValAlaArgValGlyProValSerValAlaIleAspAlaSerLeuThr 254

806 TTTTCTTCTGTAATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 862
|||||
255 SerThrGlnLysLeuTrpLysGlyValLysTrpLysAspLysAsnLysSerSerAspAsnVal 274

863 AATGATGACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 922
|||||
275 AsnHisAlaValLeuAlaValGlyLysLysLysLysLysLysLysLysLysLysLysLys 294

923 AAAAGCGCTGAGGCTCAATCTTGTGATGATGATGATGATGATGATGATGATGAT 982
|||||
295 LysAsnSerTrpGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 314

983 GGAATGACTGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1027
|||||
315 AsnAsnAlaCysGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 329

RESULT 8
US-10-114-464-7
Sequence 7, Application US/10114464
Patent No. US20020142448A1
GENERAL INFORMATION:
APPLICANT: Hastings, et al.
TITLE OF INVENTION: HUMAN OSTEOCLAST-DERIVED CATHEPSIN
FILE REFERENCE: PRI0705
CURRENT APPLICATION NUMBER: US/10/114,464
CURRENT FILING DATE: 2002-04-03
PRIOR APPLICATION NUMBER: 08/553,125
PRIOR FILING DATE: 1995-11-07
PRIOR APPLICATION NUMBER: 08/208,007
PRIOR FILING DATE: 1994-03-08
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 7
LENGTH: 329
TYPE: PRT
ORGANISM: Homo sapiens
US-10-114-464-7

Alignment Scores:

Prod. No.:	1,61c-91	Length:	329
Score:	944.50	Matches:	186
Percent Similarity:	69.55%	Conservative:	47
Best Local Similarity:	55.52%	Mismatches:	79
Query Match:	47.30%	Indels:	24
DB:	12	Gaps:	5

US-10-010-580-1 (1-1126) x US-10-114-464-7 (1-329)

54 CTGCTCCGCTGCTGCTCTATGACGTGACGACCAATACATTAAGATCCACTGATCAT 113
|||||
8 LeuLeuProValValSerPheAla-----LeuHisProGluGluIleLeuAspThr 24

114 CACTGCAATCTGCAAGAAATTTAGAGAAATTTAGAGAAATTTAGAGAAATTTAGAG 172
|||||
25 GlnTrpGluLeuTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 44

173 GCACGGCGCTGATCGGCAAAAAAATCTAAATTTGCTGATGCTGACATCTGGAAAC 232
|||||
45 SerArgGlnLeuLeuTrpGluLysAsnLeuLysHisLysSerLysHisAsnLeuGluAla 64

233 TCATGCAATGATCATATGATGATGATGATGATGATGATGATGATGATGATGAT 292
|||||
65 SerLeuGluValHisThrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 84

293 GAAGTCAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 352
|||||
85 GluValValGlnLysMetThrGlyLeuLysValProProSer-----CCTGAT 98

353 TATAGTCAAACTCTAATCAAAATG-----CCTGAT 385
|||||
99 ArgSerHisSerAsnAspThrLeuTrpLysLeuProAspTrpGluLysArgThrProAsp 117

386 TCTGTGAGCTGAG 445
|||||
118 SerLysAspTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 137

446 GCTTGGGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 505
|||||
138 SerGlySerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 157

PRIOR FILING DATE:	1997-12-11
PRIOR APPLICATION NUMBER:	60/068696
PRIOR FILING DATE:	1997-12-16
PRIOR APPLICATION NUMBER:	60/072320
PRIOR FILING DATE:	1998-01-23
PRIOR APPLICATION NUMBER:	60/073611
PRIOR FILING DATE:	1998-02-04
PRIOR APPLICATION NUMBER:	60/074086
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/074099
PRIOR FILING DATE:	1998-02-09
PRIOR APPLICATION NUMBER:	60/077799
PRIOR FILING DATE:	1998-03-12
PRIOR APPLICATION NUMBER:	60/078910
PRIOR FILING DATE:	1998-03-20
PRIOR APPLICATION NUMBER:	60/079299
PRIOR FILING DATE:	1998-03-25
PRIOR APPLICATION NUMBER:	60/079666
PRIOR FILING DATE:	1998-02-27
PRIOR APPLICATION NUMBER:	60/079728
PRIOR FILING DATE:	1998-03-27
PRIOR APPLICATION NUMBER:	60/080166
PRIOR FILING DATE:	1998-03-31
PRIOR APPLICATION NUMBER:	60/081200
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081226
PRIOR FILING DATE:	1998-04-09
PRIOR APPLICATION NUMBER:	60/081699
PRIOR FILING DATE:	1998-04-11
PRIOR APPLICATION NUMBER:	60/081811
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/081818
PRIOR FILING DATE:	1998-04-15
PRIOR APPLICATION NUMBER:	60/082999
PRIOR FILING DATE:	1998-04-24
PRIOR APPLICATION NUMBER:	60/084322
PRIOR FILING DATE:	1998-04-28
PRIOR APPLICATION NUMBER:	60/083544
PRIOR FILING DATE:	1998-04-29
PRIOR APPLICATION NUMBER:	60/084600
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/084627
PRIOR FILING DATE:	1998-05-07
PRIOR APPLICATION NUMBER:	60/084633
PRIOR FILING DATE:	1998-05-11
PRIOR APPLICATION NUMBER:	60/085149
PRIOR FILING DATE:	1998-05-12
PRIOR APPLICATION NUMBER:	60/085322
PRIOR FILING DATE:	1998-05-11
PRIOR APPLICATION NUMBER:	60/085699
PRIOR FILING DATE:	1998-05-15
PRIOR APPLICATION NUMBER:	60/085704
PRIOR FILING DATE:	1998-05-15
PRIOR APPLICATION NUMBER:	60/086414
PRIOR FILING DATE:	1998-05-22
PRIOR APPLICATION NUMBER:	60/086433
PRIOR FILING DATE:	1998-05-22
PRIOR APPLICATION NUMBER:	60/087106
PRIOR FILING DATE:	1998-05-28
PRIOR APPLICATION NUMBER:	60/088026
PRIOR FILING DATE:	1998-06-04
PRIOR APPLICATION NUMBER:	60/088737
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088744
PRIOR FILING DATE:	1998-06-10
PRIOR APPLICATION NUMBER:	60/088810
PRIOR FILING DATE:	1998-06-10

1	PRIOR APPLICATION NUMBER: 60/0088658
2	PRIOR FILING DATE: 19/98-06-11
3	PRIOR APPLICATION NUMBER: 60/00893522
4	PRIOR FILING DATE: 1998-06-17
5	PRIOR APPLICATION NUMBER: 60/00895939
6	PRIOR FILING DATE: 1998-06-17
7	PRIOR APPLICATION NUMBER: 60/0089907
8	PRIOR FILING DATE: 1998-06-18
9	PRIOR APPLICATION NUMBER: 60/0089947
10	PRIOR FILING DATE: 1998-06-19
11	PRIOR APPLICATION NUMBER: 60/090349
12	PRIOR FILING DATE: 1998-06-23
13	PRIOR APPLICATION NUMBER: 60/090429
14	PRIOR FILING DATE: 1998-06-24
15	PRIOR APPLICATION NUMBER: 60/090445
16	PRIOR FILING DATE: 1998-06-24
17	PRIOR APPLICATION NUMBER: 60/090538
18	PRIOR FILING DATE: 1998-06-24
19	PRIOR APPLICATION NUMBER: 60/090863
20	PRIOR FILING DATE: 1998-06-26
21	PRIOR APPLICATION NUMBER: 60/091360
22	PRIOR FILING DATE: 1998-07-01
23	PRIOR APPLICATION NUMBER: 60/091519
24	PRIOR FILING DATE: 1998-07-02
25	PRIOR APPLICATION NUMBER: 60/091982
26	PRIOR FILING DATE: 1998-07-07

Alignment Scores:	
Pred. No.:	9,996-81
Score:	843.00
Percent Similarity:	65.36%
Best Local Similarity:	50.60%
Query Match:	42.71%
DH:	9
	1
	5
Length:	33.3
Matches:	166
Conservative:	49
Mismatches:	106
Indels:	10
Gaps:	5

US-10-010-580-1 (1-1126) x US-10-028-072-12 (1-334)

[illegible]


```

US 10-121-904-12
? Publication No.: US20030022428A1
? GENERAL INFORMATION:
? APPLICANT: Baker, Kevin P.
? APPLICANT: Beresini, Maureen
? APPLICANT: Desnoyers, Luc
? APPLICANT: Desnoyers, Luc
? APPLICANT: Filvaroli, Ellen
? APPLICANT: Gao, Mei-Qiang
? APPLICANT: Gerlitsen, Mary E.
? APPLICANT: Goddard, Audrey
? APPLICANT: Godowski, Paul J.
? APPLICANT: Gurney, Austin L.
? APPLICANT: Sherwood, Steven
? APPLICANT: Smith, Victoria
? APPLICANT: Stewart, Timothy A.
? APPLICANT: Tumas, Daniel
? APPLICANT: Watanabe, Collin K
? APPLICANT: Wood, William
? APPLICANT: Zhang, Zemin
? TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
? FILE OF INVENTION: ACIDS ENCODING THE SAME
? FILE REFERENCE: P3330R1C54
? CURRENT APPLICATION NUMBER: US/10/123,904
? PRIOR FILING DATE: 2002-04-16
? NUMBER OF SEQ ID NOS: 550
? SEQ ID NO 12
? LENGTH: 334
? TYPE: PRT
? ORGANISM: Homo Sapien
US 10-121-904-12

Alignment Scores:
Pred. No.: 9,990-81 Length: 334
Score: 843.00 Matches: 168
Percent Similarity: 65.46% Conservative: 49
Best Local Similarity: 50.60% Mismatches: 105
Query Match: 42.21% Indels: 10
DB: 9 Gaps: 5

US 10-010-580-1 (1126) x US-10-121-904-12 (1334)
QY 54 GTGCTTGCTGTGTGCTGCTATGCGACGTGCGACAGTACATATAA---GATCCGACTCTGTGAT 110
DB / LeuAlaIalaphcysleuglyleAlaserAlavalProlyspheaspglnAsnleuasp 26
QY 111 CATCACTGCAATGCTTCGACAGCAAAACCTACACCAACATGACAGCAATGACAGAG 170
DB 27 ThrlystrpyrglntrprlyslalathrlsArgrq--LeutlyGlyAlaAsncluglug 46
QY 171 TAGCAAGGCGTCTATGTTGGCAAAAATAATTAATTGTGATATCTCAAAATCTGAAAC 230
DB 46 lytrprgrdAlavaltrprclulysAsnMetlysmclleclulclunlsAsncluglug 66
QY 241 ACTCAATGCAATGATATATATATATATATATATATATATATATATATATATATATATAT 290
DB 66 yfserglnglylsylnglylsylnglylsylnglylsylnglylsylnglylsylnglylsylng 86
QY 291 AACAACTGCAATGCTTCGACAGTTCGCCAGAGTTCGCCAGCAATGCGACAGCAATGCA 350
DB 86 lughlphrvalglmeclmetcllyksrphedq---Aasnglnlysrphedqsllylsy 105
QY 451 CTATAGCTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 410
DB 105 alpheargluprleuPhelcylusAsnleuAsnleuAsnleuAsnleuAsnleuAsnleuAsn 125
QY 411 GTGTACTGAAGTAAATGCAAGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 470
DB 125 yfvalthPrvalthlyAsnclnlysglnlysglnlysglnlysglnlysglnlysglnlysgln 145
QY 471 GGGCTCTCGAAGCAATCACTCAAGCTAAACAGCAAGCAAGCTCTCTCTCTCTCTCTCTCTCA 530

```

```

DB 145 lyAlaleucluglyclnmetrphArgllysthrcllylsleuvalSerleuSercluglnA 165
QY 531 ACTTGAGTAGTCTCTAATGAAAAATATGAAATATGAAATATGAAATATGAAATATGAAAT 590
DB 165 snleuvalAspCysSerArgrProglh---clysnclnclnglylsAsnclnglylspheme 184
QY 591 CAACGCTTTCCAGTATATATATATATATATATATATATATATATATATATATATATATAT 650
DB 184 laArgAlaPhedglntlyvalysglusncllyglleuAsnpsrsergluglnserTyrProt 204
QY 651 ACAAGCCATCAATGCAAGTGCAGATATACATCACTCAAAAAACGACGCTGCCCATGTTCAA 710
DB 204 yvalAlaValAspclulclcylysltyrArgrProglusnserValAlaAsnAspThrG 224
QY 711 AGTATACAGTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 770
DB 224 lyPhetrValAlaAlaProgllylsclulysAlaIaleucllyslAlaValAlaThrValG 244
QY 771 GACCTGCTCTGTGCTATAGATGCGACGACATATCTTCTCTCTCTCTCTCTCTCTCTCTCT 830
DB 244 lyProIleSerValAlaIaleucllysltyrArgrProglusnserValAlaAsnAspThrG 264
QY 831 TCTACTATGACATCTCTCT---ACTCAATGCAATGCAATGCAATGCAATGCAATGCAATG 887
DB 264 lytyrPhedgluprAspCysSerSerlylsAsnleuAsnphlsclyAlaIaleuValAlclY 284
QY 888 ATGCT-----AACTTAATGGAAGATATCTGCTTTGGAAGAAATAGCTGAG 935
DB 284 yrclylphedgluglyAlaAsnserAsnAsnserlystlytrprleuvallylsAsnserTyrG 304
QY 936 GCCTCAACTTGTGCTGACCAAGATATATGCGATGCGCAAGCAAGCAAGCAAGCAAGCAAG 995
DB 304 lyProglutprgllyserAsncllytyrVallyslleAlaIaleuAsnlyslAsnAsnHlscysG 324
QY 996 GCATTCGTACTATTCCTCTTACCCAGCAATG 1027
DB 324 lytleAlatrAlaIaleuSerlyrProAsnval 334

RESULT 15
US-10-140-470-12
? Sequence 12, Application US/10140470
? Publication No.: US20030022331A1
? GENERAL INFORMATION:
? APPLICANT: Baker, Kevin P.
? APPLICANT: Beresini, Maureen
? APPLICANT: Desnoyers, Luc
? APPLICANT: Desnoyers, Luc
? APPLICANT: Filvaroli, Ellen
? APPLICANT: Gao, Mei-Qiang
? APPLICANT: Gerlitsen, Mary E.
? APPLICANT: Goddard, Audrey
? APPLICANT: Godowski, Paul J.
? APPLICANT: Gurney, Austin L.
? APPLICANT: Sherwood, Steven
? APPLICANT: Smith, Victoria
? APPLICANT: Stewart, Timothy A.
? APPLICANT: Tumas, Daniel
? APPLICANT: Watanabe, Collin K
? APPLICANT: Wood, William
? APPLICANT: Zhang, Zemin
? TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
? FILE OF INVENTION: ACIDS ENCODING THE SAME
? FILE REFERENCE: P3330R1C160
? CURRENT APPLICATION NUMBER: US/10/140,470
? PRIOR FILING DATE: 2002-05-06
? NUMBER OF SEQ ID NOS: 550
? SEQ ID NO 12
? LENGTH: 334
? TYPE: PRT
? ORGANISM: Homo Sapien
US-10-140-470-12

```



```
|||||
Db 637 AlapheserAlaValGlyAlaLeuGlnLeuLysLeuLysThrGlyLeuVal 656
QY 515 TCTCTGAGTGCACAAATTTGGTAGATTGCTCAACTGAAATAATGGAAATAAAGGCTGC 574
Db 657 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlyLysGlyAsnLysGlyCys 676
QY 575 AATGGCGTTTCATGACAACTGCTTCAGATATATTATGATTAACCAACGGCATGATGCA 634
Db 677 AsnGlyGlyPheMetThrThrAlaPheGlnIleLeuAspAsnLysGlyIleAspSer 696
QY 635 GAAGCTTCCTATCCCTACAAAGCCATGCAATGCAAGTGCAGATATGACTCAAAAACGCA 694
Db 697 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 716
QY 695 GCTGGCCACATGTTCAAGTATACGTAATCTCCCTTGGCAGTGAAGATGCTTAAAGAA 754
Db 717 AlaAlaIleThrCysSerLysTyrThrGlyLeuProTyrGlyArgGlyAspValLeuLysGly 736
QY 755 GCTGTGGCCAAATAAGACCTGTGTCTGTGGCTATAGATGGACCCACTATTCTTTCTTC 814
Db 737 AlaValAlaAsnLysGlyProValSerValGlyValAspAlaArgHisProSerPhePhe 756
QY 815 CTGTACAGAACTGGTGTCTACTATGACACCATCTGTACTCGAATGTGCAATCATGAGATA 874
Db 757 LeuTyrArgSerGlyValTyrTyrGlyProSerCysThrGlnAsnValAsnHisGlyVal 776
QY 875 TTAGGGTGGCTATGGTAACTTAATGGAAAGCTGCTGGTGGAAAGAAAGAGCTGG 934
Db 777 LeuValValGlyTyrGlyAspLeuAsnLysGlyLeuTyrTrpLeuValLysAsnSerTyr 796
QY 935 GGCCTCAACTTGTGTGACCAAGATATATTCGGATGGCAGAAACAGTGGAAATCACTGT 994
Db 797 GlyHisAsnPheGlyGlnGlyGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 816
QY 995 GGGATGCTGCTATTCCTCTTACCCAGAAATC 1027
Db 817 GlyIleAlaSerPheProSerTyrProIle 827

RESULT 6
US-08-553-125-8
; Sequence 8, Application US/08553125
; GENERAL INFORMATION:
; APPLICANT: HASTINGS, ET AL.
; TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BRCKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,125
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/208,007
; FILING DATE: March 8, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRARO, GREGORY D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-505
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
```

```
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; US-08-553-125-8

Alignment Scores:
Pred. No.: 2,27e-157 Length: 331
Score: 1569.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.57% Indels: 1
DB: Gaps: 0

US-10-010-580-1 (1-1126) x US-08-553-125-8 (1-331)
QY 36 ATGAATGCTAGTGGCTGCTTCGCTTCCTATGCACTGGCAGCAATGATAA 95
Db 1 MetLysArgLeuValCysValLeuLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
QY 96 GATCCCACTGTGATCATCACTGGAATCTCTGCAAGAAAACCTACAGCAAAATC-AAG 154
Db 21 AspProThrLeuAspHisHisIleThrHisLeuTrpLysLysThrTyrGlyLysGlnTyrLys 40
QY 155 GAAAGAGATAGAGAAAGACAGCAGCGCTCATATCTGGAAAAAATTTAAATTTGTATG 214
Db 41 GlyLysAsnGlnGlyAlaValArgArgLeuIleTrpGlnLysAsnLeuLysPheValMet 60
QY 215 CTTCACAAATGTGAACACCTCAATGGAATGATCATGATGATGATGATGCAAGCAATCTG 274
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTyrAspLeuIleMetAsnHisLeu 80
QY 275 GGAACATGACCTGGTGAAGAAGTATATCTTATGAGGTTCCTGACAGCTCCACAGCAA 334
Db 81 GlyAspMetThrSerGlnGlyValMetSerLeuMetSerSerLeuArgValProSerGln 100
QY 335 TGGCAGAGAAATGTCACCTTATAGTCAAACTTAATCATGAAATTTGGCTCATTTGTGAC 394
Db 101 TrpGlnArgAsnIleThrTyrTyrLysSerAspProAsnArgIleLeuProAspSerValAsp 120
QY 395 TGGAGAGAAAGAGCGCTGTCTACTGAAGTGAATATACCAGCGTCTGTGCTGCTGTTCG 454
Db 121 TrpArgGlnLysGlyCysValThrGlnValLysTyrGlnGlySerCysGlyAlaCysTrp 140
QY 455 GCTTTCAGCGCTGTGGGGCCCTGGAAAGCAACTGAAGCTAAAGAACAGAAAGCTGTG 514
Db 141 AlapheserAlaValGlyAlaLeuGlnLeuGlnLeuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTGTGATGTCACAGAACTGTAGATTGCTCAACTGAAATAATGGAATAAAGCTGAC 574
Db 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysTyrGlnLysAsnLysGlyCys 180
QY 575 AATGGCGTTTCATGACAACTGCTTCAGATATATTATGATTAACCAACGGCATGATGCA 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnIleLeuAspAsnLysGlyIleAspSer 200
QY 635 GAAGCTTCCTATCCCTACAAAGCCATGCAATGCAAGTGCAGATATGACTCAAAAACGCA 694
Db 201 AspAlaSerTyrProTyrLysAlaMetAspLeuLysCysGlnTyrAspSerLysTyrArg 220
QY 695 GCTGGCCACATGTTCAAGTATACGTAATCTCCCTTGGCAGTGAAGATGCTTAAAGAA 754
Db 221 AlaAlaIleThrCysSerLysTyrThrGlnLeuProTyrGlyArgGlnAspValLeuLysGly 240
QY 755 GCTGTGGCCAAATAAGACCTGTGTCTGTGGCTATAGATGCGAGCCACTATTCTTTCTTC 814
Db 241 AlaValAlaAsnLysGlyProValSerValGlyValAspAlaArgHisProSerPhePhe 260
QY 815 CTGTACAGAGCTGTGCTACTATGAAACATCTGTACACAGAAATGTAATCATGACAGTA 874
```

Db 261 LeuTYrArGserGIyValTYrGLuProSerCysThrGlnAsnValAsnHisGLyVal 280
QY 875 TTATGGTTGGTATGATACCTTAATGGCAAAACACTGTGGCTTTGTAAATAACAGCTGG 934
Db 281 LeuValValGlyTYrGlyValAspLeuAsnGlyLysGlyTYrTrpLeuValLysAsnSerTrp 300
QY 945 GCGCTCAACCTTTGGTGGACCAAGATATATTCGATGGCAAGAACACTGTCAAACTCAGT 994
Db 301 GlyHisAsnPhenGlyGluGlyLysGlyTYrLeuArgMetAlaArgAsnLysGlyAsnHisGly 320
QY 995 GGGATTGCTATTTATCCCTCTTACCGCAAAATC 1027
Db 321 GlyIleAlaSerPheProSerTYrProGluIle 331

RESULT 7
US 09 219-441-B
Sequence H, Application US/09219441
GENERAL INFORMATION:
APPLICANT: HASTINGS, ET AL.
TITLE OF INVENTION: Human Osteoclast-Derived Cathepsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 HECKER PARK ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/219,441
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/208,007
FILING DATE: March 8, 1994
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/EXAMINER NUMBER: 325800-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 994-1700
TELEFAX: 201 994-1744
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLACT: LINEAR
MOLECULE TYPE: PROTEIN
US 09 219-441-B

Alignment Scores:
Prod. No.: 2-27e-157 Length: 331
Score: 1569.00 Matches: 290
Percent Similarity: 93.15% Conservatve: 19
Host Local Similarity: 87.61% Mismatches: 22
Query Match: 78.57% Indels: 1
DB: 16 Gaps: 0

US 10-010-580-1 (1 1126) x US-09-219-441-B (1-331)
QY 46 ATGAATGGCTACTGGGCTGGCTGGCTGGCTATGCGAGTGGCAAGTACATATAA 95
Db 1 MetLysArgLeuValGlyValLeuValGlySerSerAlaValAlaGlnLeuHisLys 20
QY 96 GATTCGATCTTCGATCATCATGCAATCTCTGGAGAAACCTTACGCAACATTC-AAC 154
|||||
|||||

Db 21 AsperothrLeuAsnPhHisHisThrPheLysLeuTrpLysLysThrTYrGlyLysGlnTYrLys 40
QY 155 GAAGAGATATAGAGAAATAGATGGGCTTTATTTAGAAAAAAATTTAAATTTGTGATG 214
Db 41 GluLysAsnGluGluAlaValArgArgLeuIleTrpGluLysAsnLeuLysPheValMet 60
QY 215 CTTTCACATATGGAGACAGCTCAATGGGAATGCATTCATGATGATGATGATGATGATG 274
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTYrAspLeuGlyMetAlaHisLys 80
QY 275 GGAACATGATGCTGTGAAGAAGTATATCTTGATGGGATTCCTGAGATTCGAGCCAA 334
Db 81 GlyAspMetThrSerGluGluValMetSerLeuMetSerSerLeuArgValProSerGln 100
QY 335 TGGCAGAGAAATGCTATTTAGTCAAAATTTATACAGAAATTTGCTTATCTGTGGAC 394
Db 101 TrpGlnArgAsnIleThrTYrLysSerAsnProAsnArgIleLeuProAspSerValAsp 120
QY 395 TCGACAGAGAGGCGCTGTGTACTGAAGTCAAAATACAGGCTTCGTGGCTGGTGG 454
Db 121 TrpArgGluLysGlyValThrGluValLysTYrGlnGlySerCysGlyAlaCysTrp 140
QY 455 GCTTCAGCGCTGTGGGGGCGCTGGAAACACAACTAAGCTAAACAGAGAAAGCTGGT 514
Db 141 AlaPheSerAlaValGlyAlaLeuGluAlaGlnLeuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTGTAGTGCACAGAACTGGTAGATTGCTCACTGAATAATATGGAATAAGGCTGC 574
Db 161 SerLeuSerIleGlnAsnLeuValAspCysSerThrGluLysTYrGlyLysLysLys 180
QY 575 AATGGCGGTTTCATGACAACTGCTTCCACTATATATTGATTAACACGCAATTCATGCA 634
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnTYrIleAlaAspAsnLysGlyIleAspSer 200
QY 635 GAAGCTTCATCCCTACAAAGCCTGAGTGAATGGAAATGGAATATATGCAAAAAAGCA 694
Db 201 AspAlaSerTYrProTYrLysAlaMetAspLeuLysCysGlnTYrAspSerTYrArg 220
QY 695 GTTGCACATGTTCAAAATATACATACTTCCTTGGAGTGAAGATGCTTAAAGAA 754
Db 221 AlaAlaThrCysSerLysTYrThrGluLeuProTYrGlyArgGluAspValLeuGlu 240
QY 755 GCTGTGGCAATTAAGACCTGTGCTGTGGCTATAGATGCCAACACATCTTCTTC 814
Db 241 AlaValAlaLysLysGlyProValSerValGlyValAspAlaArgHisProSerPhePhe 260
QY 815 CTGTACAGAAAGTGGTGTACTACTATGAAACCATCTGTACTCAGAAATGTAATCATGAGTA 874
Db 261 LeuTYrArgSerGlyValTYrGluProSerCysThrGlnAsnValAsnHisGLyVal 280
QY 875 TTATGGTTGGTATGATACCTTAATGGCAAAACACTGTGGCTTTGTAAATAACAGCTGG 934
Db 281 LeuValValGlyTYrGlyValAspLeuAsnGlyLysGlyTYrTrpLeuValLysAsnSerTrp 300
QY 935 GCGCTCAACCTTTGGTGGACCAAGATATATTCGATGGCAAGAACACTGTCAAACTCAGT 994
Db 301 GlyHisAsnPhenGlyGluGlyLysGlyTYrLeuArgMetAlaArgAsnLysGlyAsnHisGly 320
QY 995 GGGATTGCTATTTATCCCTCTTACCGCAAAATC 1027
Db 321 GlyIleAlaSerPheProSerTYrProGluIle 331

RESULT 8
US-09-791-537-94624
Sequence 94624, Application US/09/91537
GENERAL INFORMATION:
APPLICANT: Bionomix, Inc.
APPLICANT: Debe, Derek
APPLICANT: Dauter, Joseph
TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
FILE REFERENCE: 261/210
CURRENT APPLICATION NUMBER: US/09/791,537


```

Db      141  Alahhseralalavalalclgylalaleuglualaglnleuylsleuylthrclylsleuval 160
QY      515  TCTGTGAGTGTATATGAACTTGATGATTTGCTCAACTGAGAAAAATATGGAAATTAAGGCTGC 578
Db      161  ThrlleuseralaglnaslnleuvalAspCysSerThrGluLysTyrClyAsnLysclYcys 180
QY      575  AATGCGCGTTTCATGACAACTGCGTTTCGACGTATATTATTCATGAACAACGCGCATTCATCA 634
Db      181  AsnclgylPhemeLthrThrAlaPhneGlnTyrIlelleAspAsnLysGlyLeAspSer 200
QY      635  CAACCTTCCTCATCGCTTCAACAAAGCCATGATGCAAAAGTCAGACATATCACTCAAAAAAGCGA 694
Db      201  AspalAserLyrProTyrLysAlaMetLaspGlnLysCysGlnTyrAspSerLysTyrArg 220
QY      695  GTGTGTCACATTTCAAAAGTATACCTGAATCTCCATTGGACATGGAAGATGCGCTTAAAGAA 754
Db      221  AlalalathrcysSerLysTyrThrGluLeuProTyrGlyArgGlyAspValLeuLysGlu 240
QY      755  CCTGTGCCCAATTAAGACGCTGTGTCTGTGCGCTATATAGATGCGAGCCACTATCTTTCTTC 814
Db      241  AlalalalAsnLysClyProvalSerValGlyValAspAlaArgHisProSerPhe 260
QY      815  CTCGTACGACAGTGGTGTCTATCTATGAACCATCGCTGTACACAAATGCAATCATCGAGTA 874
Db      261  LeuLyrArgSerGlyAlaLysTyrGlyLupProSerCysThrGlnAsnValAsnHisGlyVal 280
QY      875  TTACTGTGTAATATGATGTAACCTTAAATGGAAGAAGTACTAGCTGTGTGAAAAAGAGCTGG 934
Db      281  LeuValValGlyTyrClyAspLeuAsnGlyLysGluTyrThrLeuValLysAsnSerTrp 300
QY      935  GCGCTCAACCTTGGCTGACCAAGCATATATTCGATGCGCAAGAAACAGTGCAAATCAGCTGT 994
Db      301  GLyHisAsnPhneGlyGlnGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
QY      995  GCGATTCCTACTTATTCGCTCCCTTAAACCCAGAAATC 1027
Db      321  GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 13
US-09-463-000-2
: Sequence 2, Application US/09463000
: GENERAL INFORMATION:
: APPLICANT: Abdel-Meguid, Sherin
: APPLICANT: Janson, Cheryl
: APPLICANT: Lalonde, JudyLh
: APPLICANT: Smith, Ward
: APPLICANT: Zhao, Baoquang
: TITLE OR INVENTION: Method of inhibiting Cathepsin K
: FILE REFERENCE: P50684
: CURRENT APPLICATION NUMBER: US/09/463,000
: CURRENT FILING DATE: 2000-01-14
: PRIOR APPLICATION NUMBER: 60/052,831
: PRIOR FILING DATE: 1997-07-17
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FASTSQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapien
US-09-463-000-2

Alignment Scores:
Pred. No.: 1,26e-156 Length: 331
Score: 1562.00 Matches: 288
Percent Similarity: 93.05% Conservative: 20
Best Local Similarity: 87.01% Mismatches: 23
Query Match: 78.42% Indels: 1
DB: 18 Gaps: 0

US-10-010-580-1 (1-1126) x US-09-463-000-2 (1-331)
36 ATCAAAATGCGTACTTGCGCGCTCGCTCCGTTGCTCTCCATATGACAGTGGCAAGATGACATPAA 95

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: June 27, 2003, 11:29:39 ; Search time 133 Seconds
(Without alignments)
4393.190 Million cell updates/sec

Title: US-10-010-580-1

Perfect score: 1997

Sequence: 1 gcatatagatagagagca.....aataatgtagatgacat 1126

Scoring table:

HiOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1184592 seqs, 259455700 residues

Total number of hits satisfying chosen parameters: 2369184

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame.n2p.model -DEV=xip
-O/cgn2_1/USPTO.spool/US10010580/runat_27062003.104538.11821/app_query.fasta_1.1287
-DB-pending_Patents_AA_New -QFMT-fastan -SUFFIX-n2p.rapn -MINMATCH-0.1
-LOOPCL-0 -LOOPEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-biosum62
-TRANS-human40.cdl -List-45 -DOCALLIGN-200 -THR SCORE-pct -THR MAX-100
-THR MIN-0 -ALIGN-15 -MODE-LOCAL -OUTFMT-pct -NOR-ext -HEARSTIE-500 -MINLEN-0
-NCPU-6 -ICPU-3 -NO_MMAP -LAKEROBERT -NEG_SCORES-0 -WAIT -DSPBLOCK-100
-LONGLOG -DEV-TIMEOUT-120 -WARN-TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5
-Fgapop-6 -Fgapext-7 -Ygapop-10 -Ygapext-0.5 -DELop-6 -DELext-7

Database : Pending Patents AA New:

1: /cgn2_6/pdata/1/paa/US06_NEW_COMB.pep.*
2: /cgn2_6/pdata/1/paa/US07_NEW_COMB.pep.*
3: /cgn2_6/pdata/1/paa/US08_NEW_COMB.pep.*
4: /cgn2_6/pdata/1/paa/US09_NEW_COMB.pep.*
5: /cgn2_6/pdata/1/paa/US10_NEW_COMB.pep.*
6: /cgn2_6/pdata/1/paa/US60_NEW_COMB.pep.*
7: /cgn2_6/pdata/1/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1581.5	79.2	353	US-10-227-425-1771 Sequence 1771, Ap
2	1581.5	79.2	353	US-10-227-425-2997 Sequence 2997, Ap
3	1571	78.7	331	US-09-724-676-81732 Sequence 81732, A
4	1571	78.7	331	US-09-724-676-81733 Sequence 81733, A
5	1571	78.7	331	US-09-724-676-81732 Sequence 81732, A
6	1571	78.7	331	US-09-724-676-81733 Sequence 81733, A
7	1571	78.7	331	US-60-452-680-23324 Sequence 23324, A
8	1571	78.7	331	US-60-453-135-14611 Sequence 14611, A
9	1571	78.7	331	US-60-453-050-14611 Sequence 14611, A
10	1571	78.7	331	US-60-455-444-7889 Sequence 7889, Ap
11	1571	78.7	331	US-60-465-241-7889 Sequence 7889, Ap

12	1571	78.7	331	US-60-466-412-14611 Sequence 14611, A
13	1568	78.5	331	US-10-219-051B-2757 Sequence 2757, Ap
14	1568	78.5	331	US-10-219-051B-10849 Sequence 10849, A
15	1565	78.4	331	US-10-273-577-1 Sequence 1, Appl
16	1563	78.3	331	US-10-273-577-6 Sequence 6, Appl
17	1560	78.1	331	US-10-318-584-4 Sequence 4, Appl
18	1555	77.9	331	US-10-273-577-8 Sequence 8, Appl
19	1553	77.8	331	US-10-273-577-7 Sequence 7, Appl
20	1334.5	66.8	330	US-10-219-051B-2755 Sequence 2755, Ap
21	1334.5	66.8	330	US-10-219-051B-10847 Sequence 10847, A
22	1085	54.3	217	US-10-273-577-5 Sequence 5, Appl
23	1083	54.2	217	US-10-273-577-2 Sequence 2, Appl
24	1075	53.8	217	US-10-273-577-4 Sequence 4, Appl
25	1073	53.7	217	US-10-273-577-3 Sequence 3, Appl
26	944.5	47.3	329	US-10-318-584-3 Sequence 3, Appl
27	939.5	47.0	329	US-10-318-584-2 Sequence 2, Appl
28	939.5	47.0	329	US-10-219-051B-6306 Sequence 6306, Ap
29	939.5	47.0	329	US-10-219-051B-11531 Sequence 11531, A
30	939.5	46.9	329	US-60-474-733-201 Sequence 201, App
31	937.5	46.9	329	US-10-219-051B-6304 Sequence 6304, Ap
32	937.5	46.9	329	US-10-219-051B-11529 Sequence 11529, Ap
33	934.5	46.8	329	US-60-452-680-23323 Sequence 23323, A
34	934.5	46.8	329	US-60-453-135-14610 Sequence 14610, A
35	934.5	46.8	329	US-60-453-135-14610 Sequence 14610, A
36	934.5	46.8	329	US-60-455-444-7888 Sequence 7888, Ap
37	934.5	46.8	329	US-60-465-241-7888 Sequence 14610, A
38	934.5	46.8	329	US-60-466-412-14610 Sequence 14610, A
39	926.5	46.4	329	US-60-474-733-7 Sequence 7, Appl
40	875	43.8	365	US-10-405-027-3738 Sequence 3738, Ap
41	843	42.2	334	US-10-131-819A-12 Sequence 12, Appl
42	843	42.2	334	US-10-131-819A-12 Sequence 12, Appl
43	843	42.2	334	US-10-131-823A-12 Sequence 12, Appl
44	843	42.2	334	US-10-131-824A-12 Sequence 12, Appl
45	843	42.2	334	US-10-131-826A-12 Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-10-227-425-1771
Sequence 1771, Application US/10227425
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT249CIN
CURRENT APPLICATION NUMBER: US/10/227,425
PRIOR FILING DATE: 2002-08-26
PRIOR APPLICATION NUMBER: 09/760,475
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/218,290
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 4122
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1771
LENGTH: 353
TYPE: PRT


```

Db      72  LeuiletrpclulysasnleuylspheValMetLeuHisnleugluHisSerMetGly 91
OY      242 ATGCATTATATATATCTAGCATGAACCATCTGGAGACATGCTGTGACAGATGATA 301
      92  MetHisSerIyrAspLeuGluMetAsnHisLeuGlyAspMetHisSerGluGluValMet 111
OY      302 TCTTTGATGGTCCCTGAGAGTCCAGCCCAATGGCAGAAATGCACTTATAGTCA 361
      112 SerLeuMetSerSerLeuArGValProSerGlnThrPheLysArgAsnHisLeuIyrLysSer 131
OY      362 AACTTATACAGAAATTCCTGATTTCTGTGAGTGGAGAGAGAGAGGCTGTGTACTGAA 421
      132 AspProAsnArGileuProAspSerValAspIrrArGluLysGlyValThrGlu 151
OY      422 GTCAATACACAGGCTTCTGTGCTGTGTGGCTTTCAGCCCTGGGGCCCGCCGAA 481
      152 ValLysIyrGlnLysSerGlyAlaCysIrrPheSerAlaValGlyAlaLeuGlu 171
OY      482 GCACAACTGAAGCTAAACAGAGAAAGCTGTCTCTGAGTGCACAGAACTTGTAGAT 541
      172 AlaGlnLeuLysLeuLysThrGlyLysLeuValSerLeuSerAlaGlnAsnLeuValAsp 191
OY      542 TGTCAACTGAAAAATATGGGATTAAGGCTGCATGGGGGTTTCATGACACTGCTTC 601
      192 CysSerThrGluLysIyrGlyAsnLysGlyCysAsnGlyLysPheMetThrThrAlaPhe 211
Db      602 CAGTATTATTATGATTAACAACAGCATGATTCAGAACTCCATCCCTACAAAGCCATG 661
      212 GlnIyrIleLeuAspAsnLysGlyLeuAspSerAlaSerIyrProIyrLysAlaMet 231
OY      662 AATGGAAGTGCAGATATGACTAAAAAAGCAGCTGCCACATGTTCAAAATATACTGAA 721
      232 AspGlnLysCysGlnIyrAspSerLysIyrArGAlaAlaThrCysSerIyrIyrThrGlu 251
OY      722 CTTCCTTTGGCAGTGAAGATGCTTAAAGAGCTGTGGCCATTAAGAGACTGTGCT 781
      252 LeuProIyrGlyArGluAspValLeuLysGluAlaValAlaAsnLysGlyIyrProValSer 271
OY      782 GTGGCTATAGATCGCAGCCACTATTCTTCTTCTGTACAGAAAGTGGTGTACTGAA 841
      272 ValGlyValAspAlaArGHisIrrProSerPhePheLeuIyrArGSerLysValIyrGlu 291
OY      842 CCACTCTGACTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 901
      292 ProSerCysThrGlnAsnValAsnHisGlyValLeuValValGlyIyrGlyAspLeuAsn 311
Db      902 GGGAAAGACTACTGGCTGTGAAAAACAGCTGGGCTCAACTTGTGACCAAGCATAT 961
      312 GlyLysGluIyrTrpLeuValLysAsnSerIrrPheLysAsnPhGlyGluGluGlyIyr 331
OY      962 ATTGGGATGGCAGAAACAGTGAATCACTGTGGGATTCGTATCCCTTACCCA 1021
      332 IleArMetAlaArGAsnLysGlyAsnHisCysGlyIleAlaSerPheProSerIyrPro 351
OY      1022 GAATTC 1027
      352 GluIle 353
Db

```

RESULT 3

```

US-09-724-676-81732
: Sequence 81732, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 81732
: LENGTH: 331
: TYPE: PRN

```

```

: ORGANISM: Homo sapiens
US-09-724-676-81732
Alignment Scores:
Pred. No.: 1,19e-150
Score: 1571.00
Percent Similarity: 93.35%
Best Local Similarity: 87.61%
Query Match: 78.67%
DB: 5
Gaps: 0

```

US-10-010-580-1 (1-1126) x US-09-724-676-81732 (1-331)

```

OY      36  ATGAATAGGCTAGTGGGCTGCTCCGTTGGTCCCTATGACAGTGCACAAATGACATAA 95
      1  MetLysArGLeuValCysValLeuLeuValCysSerSerAlaValAlaGlnLeuLys 20
OY      96  GATCCCACTCTGATCATCATCTGAAATCTCTGAAAGAAACCTACAGCAACATC-AAG 154
      21  AspProThrLeuAspHisIrrPheIrrPheLysIrrPheLysIrrPheLysIrrPheLys 40
Db      155  GAAGCAATAGCAAGTACACAGCGCTCATCTGGGAAAAAATCTAAATTTGTGATG 214
      41  GluLysAsnGluGluAlaValArGArGLeuIrrPheLysAsnLeuLysPheValMet 60
OY      215  CTTCACAACTCTGAAACACTCAATGGAAATGCATTCATATGATCTGACATGACCATCTG 274
      61  LeuHisAsnLeuGluHisSerMetCylMetHisSerIyrArPheLysGlyMetAsnHisSer 80
OY      275  GGAGCATGATGCTGGTGAAGAAAGTATATCTTTGATGGTCCCTGAGATCCACGCCAA 334
      81  GlyAspMetThrSerGluGluValMetSerLeuMetSerLeuArGValProSerGln 100
OY      335  TGGCAGCAAAATGCTACTTATAGTCAAACTCAATGCAAAATTCGATTCGATTCGTGAC 394
      101  TrpGlnArGAsnIleThrIyrLysSerAsnProAsnArGileuProAspSerValAsp 120
OY      395  TGGAGAGAGAGGGCTGTGTTACTGAAATGCAATACAGGGCTCTGTGCTGTGTTGG 454
      121  TrpArGluLysGlyCysValThrGluValLysIyrGlnCylSerGlyAlaCysIrrP 140
Db      455  GCTTCAGAGGCTGTGGGGCCCTGGAGACATCAACTGAAAGTAAATTCGAAAGTGTG 514
      141  AlaPheSerAlaValGlyAlaLeuGluAlaGlnLeuLysLeuLysThrGlyLysLeuVal 160
OY      515  TCTCTGACTGCACAGAACTGTGTAGATTGCTCAACTGCAAAATATGCAATTAAGCCATGC 574
      161  SerLeuSerAlaGlnAsnLeuValAspCysSerThrGluLysIyrGlyAsnLysGlyCys 180
OY      575  AATGGCGGTTTCATGACAACCTGTTTGCAGATATATTATGATTAACAGGCATGATTCA 634
      181  AsnGlyGlyPheMetThrThrAlaPheGlnIyrIleLeuAspAsnLysGlyIleAspSer 200
OY      635  GAAGCTTCCTATCCCTACAAAGCCATGAATGGAAGATGCAATATGATCAAAAAAGCA 694
      201  AspAlaSerIyrProIyrLysAlaMetAspGlnLysCysGlnIyrAspSerLysIyrArG 220
Db      695  GCTGCCAATGTTCAAAAGTATAGTGAACCTTCCCTTGGCAGTGAAGATGCTTAAAGAA 754
      221  AlaAlaThrCysSerLysIyrThrGlnLeuProIyrGlyArGluAspValLeuLysGlu 240
OY      755  GGTGGGCAATTAAGGACGTGTGTGTGATTAAGTATGATGATGATGATGATGATGATG 814
      241  AlaValAlaAsnLysGlyProValSerValGlyAlaAspAlaArGHisProSerPhePhe 260
OY      815  CTGTACAGAAAGTGTGTCTACTATGACCACTCTGTACTGAGATGTAATGTAATCATGATA 874
      261  LeuIyrArGSerGlyValIyrIyrGluProSerCysThrGlnAsnValAsnHisGlyVal 280
OY      875  TTATGCTGTTGGTATAGTAACTTAATGCAAAAGCACTACTGCTTGTGCAAAACAGCTGC 934
      281  LeuValValGlyIyrGlyAspLeuAsnGlyLysGlnIyrTrpLeuValLysAsnSerIrrP 300
Db

```

```

OY      94  GGCGCTCAATTTGCTTACAGCAATATATGCGATGCGCAACAACAGTGGAAATCACTGT 994
Db      101  GYHISASnheciylucluclyllyrllcagmclalaaGAsnlySGIYASnHiscys 320

OY      995  GGCGATTGCTAGCTTATTCCTCTTACCCGACAATTC 1027
Db      321  GYlIeAlaserPheproSerTyPProGluIle 331

RESULT 4
US-09-724-676-8173
: Sequence 8173, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181..4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 81733
: LENGTH: 341
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676-8173

```

```

QY 275 GGAGACATGCTGCTGAGCAAGTGAATATCTTGATGGCTCCCTGAGAGTTCCAGCCAA 334
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 81 GlyAspMetThrSerGluGluValMetSerLeuMetSerLeuArgValProSerGln 100
QY 335 TGGCAGAGAAATGTAAGTCTATAGTCAAACTCTAATCAAGAAATTCCTGATTCGTGGAC 394
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 101 TTPGlnArgAsn1LeThrTyrLysSerAsnProAsnArg1LeLeuProAspSerVal1Asp 120
QY 395 TGGAGAGAGAAAGGCTGTGTTACTGAGAGAAATCCAGGGCTCTTGGGCTGTGG 454
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 121 TTPArgGluLysGlyCysValThrGluValLysTyrGlnGlySerCysGlyValAcysTyr 140
QY 455 GCTTCACGCTGTGCGGGCCCTGAGACCAACTGAAGCTAAAAACAGAAAGCTGTG 514
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 141 AlaPheSerAlaValGlyValLeuGlnLysLeuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTGAGTGCACACACTGGTATGATGCTCACTGAAATATGGAATATGGAATATGAGCTGC 574
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysTyrGlyAsnLysGlyCys 180
QY 575 AATGGCGGTTTCATGACAACTGCTTCCAGTATATATATGATTAACACGCGATGATGCA 634
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrLeileAspAsnLysGly1LeAspSer 200
QY 635 CAAGCTTCTATCCCTACAAAGCCATGATGAAAGTCCACATATGACTCAAAAAACGCA 694
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 201 AspAlaSerTyrProTyrLysAlaMetLaspGlnLysCysGlnTyrAspSerLysTyrArg 220
QY 695 GCTGCCACATGTTCAAACTGATGACTGAACTCCCTTGGCAGTGAAGATGCTTAAAGAA 754
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 221 AlaAlaThrCysSerLysTyrThrClnLeuProTyrGlyArgGlnLysValLeuLysGln 240
QY 755 GCTGAGGCAATTAAGACCTGTGTCTGTGCTATAGATGAGACCATATTTCTTCTTC 814
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 241 AlaValAlaAsnLysGlyProValSerValGlyValAspAlaArgHisProSerPhePhe 260
QY 815 CTGACAGAAAGTGTGCTACTATGAAACATCTGCTACTGCAATGCAATGCAATGCAATG 874
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 261 LeuTyrArgSerGlyValTyrTyrGlnProSerCysThrGlnAsnValAsnHisGlyVal 280
QY 875 TTAGTGGTGGCTATGTAACCTTAATGGAAAGACTAGCTGTGTGAAAAACAGCTGG 934
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGlnTyrTrpLeuValLysAsnSerTrp 300
QY 935 GGCTCAACTTTGGTGGACCAAGGATATATTGGAGTGCAGAAACAGTGGAAATCATCTGT 994
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 301 GlyHisAsnPheGlyGlnGlyTyrTyrLeuArgMetAlaArgAsnLysGlyAsnHisCys 320
QY 995 GCGATTGCTAGTATCCCTCTTACCCAGAAATC 1027
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 321 GlyIleAlaSerPheProSerTyrProGluIle 331

```

RESULT 6

```

US-09-724-676A-81733
: Sequence 81733, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724,676A
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 81733
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676A-81733

```

Alignment Scores:

```

Pred. No.: 1,19e-150 Length: 331
Score: 1571.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19

```

```

Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.67% Indels: 1
DB: 5 Gaps: 0
US-10-010-580-1 (1-1126) x US-09-724-676A-81733 (1-331)
QY 36 ATGAAAGGCTACTGGGCTGCTCCCTGCTGCTATGCACTGAGCAGCAAGTACATAA 95
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 MetLysArgLeuValCysValLeuLeuValCysSerSer1AlaValAlaGlnLeuHisLys 20
QY 96 GATCCACCTGATCATCTGATGATCTGTGGAAGAAACCTACAGCAAAACATC-AG 154
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 21 AsprothrLeuAsnPhsHisThrHisLeuThrLysLysThrTyrGlyLysGlnTyrLys 40
QY 155 GAAGAGATAGAGAGATAGCAGCGGCTGATCTGGGAAAAAATCTAAATTTGTGATG 214
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 41 GlnLysAsnGlnGlnLysAlaValArgArgLeuIleTrpLysAsnLeuLysPheValMet 60
QY 215 CTTCACAAATGTCGAACACTGAATGGGAATGCAATTCATGATGATGATGATGATGATGATG 274
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTyrAspLeuGlyMetLysHisLeu 80
QY 275 GGAGACATGACTGCTGGAAGATGATATCTTGAATGGGTTCCCTGAGAGTTCCAGCCAA 334
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 81 GlyAspMetThrSerGluGluValMetSerLeuMetSerLeuArgVal1ProSerGln 100
QY 335 TGGCAGAGAAATGTAAGTCTATAGTCAAACTCTAATCAAGAAATTCCTGATTCGTGGAC 394
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 101 TTPGlnArgAsn1LeThrTyrLysSerAsnProAsnArg1LeLeuProAspSerVal1Asp 120
QY 395 TGGAGAGAGAAAGGCTGTGTTACTGAGAGAAATCCAGGGCTCTTGGGCTGTGG 454
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 121 TTPArgGluLysGlyCysValThrGluValLysTyrGlnGlySerCysGlyValAcysTyr 140
QY 455 GCTTCACGCTGTGCGGGCCCTGAGACCAACTGAAGCTAAAAACAGAAAGCTGTG 514
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 141 AlaPheSerAlaValGlyValLeuGlnLysLeuLysLeuLysThrGlyLysLeuVal 160
QY 515 TCTGAGTGCACACACTGGTATGATGCTCACTGAAATATGGAATATGGAATATGAGCTGC 574
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlnLysTyrGlnLysGln 180
QY 575 AATGGCGGTTTCATGACAACTGCTTCCAGTATATATGATTAACACGCGATGATGCA 634
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrLeileAspAsnLysGly1LeAspSer 200
QY 635 GAAGCTTCTATCCCTACAAAGCCATGATGAAAGTGCAGATATGACTCAAAAAACGCA 694
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 201 AspAlaSerTyrProTyrLysAlaMetLaspGlnLysCysGlnTyrAspSerLysTyrArg 220
QY 695 GCTGCCACATGTTCAAACTGATGACTGAACTCCCTTGGCAGTGAAGATGCTTAAAGAA 754
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 221 AlaAlaThrCysSerLysTyrThrClnLeuProTyrGlyArgGlnLysValLeuLysGln 240
QY 755 GCTGAGGCAATTAAGACCTGTGTCTGTGCTATAGATGAGACCATATTTCTTCTTCTTC 814
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 241 AlaValAlaAsnLysGlyProValSerValGlyValAspAlaArgHisProSerPhePhe 260
QY 815 CTGACAGAAAGTGTGCTACTATGAAACATCTGCTACTGCAATGCAATGCAATGCAATG 874
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 261 LeuTyrArgSerGlyValTyrTyrGlnProSerCysThrGlnAsnValAsnHisGlyVal 280
QY 875 TTAGTGGTGGCTATGTAACCTTAATGGAAAGACTAGCTGTGTGAAAAACAGCTGG 934
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGlnTyrTrpLeuValLysAsnSerTrp 300
QY 935 GGCTCAACTTTGGTGGACCAAGGATATATTGGAGTGCAGAAACAGTGGAAATCATCTGT 994
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 301 GlyHisAsnPheGlyGlnGlyTyrTyrLeuArgMetAlaArgAsnLysGlyAsnHisCys 320
QY 995 GCGATTGCTAGTATCCCTCTTACCCAGAAATC 1027
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 321 GlyIleAlaSerPheProSerTyrProGluIle 331

```

```
RESULT 7
US-60-452-680-23324
: Sequence 23324, Application US/60452680
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: GROPE, Andrew
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001450
: CURRENT APPLICATION NUMBER: US/60/452,680
: NUMBER OF SEQ ID NOS: 2003-03-07
: SOFTWARE: FASTSEQ for Windows version 4.0
: SEQ ID NO 23324
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-452-680-23324

Alignment Scores:
Pred. No.: 1,19e-150 Length: 331
Score: 1571.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.67% Indels: 1
DB: Gaps: 0

US-10-010-580-1 (1-1126) x US-60-452-680-23324 (1-331)

OY 36 ATGAATGGCTAGTGGGCTGCTCCCTGTCCTGCTATGCGGCAAGTACATATA 95
Db 1 MetLysArgLeuValCysValLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
OY 96 GATCCGACATCTGATCATCATCTGAAATCTGTGAAAGAAACCTACAGCAATATC-AAG 154
Db 21 AsprProthLeuAspHisHisIstPrHisLeuTrpHisLysTrpHisLysGlnTrpLys 40
OY 155 GAGACGATGAGGAGGATGACGAGGAGTCTATCTGGGAAAAAATCTMAAATTTGTATG 214
Db 41 GlnLysAsnGlnGlnAlaValAlaArgLeuLeuTrpGlnLysAsnLeuLysPheValMet 60
OY 215 CTTCACAACTTCGCAACACTGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 274
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTrpAspLeuGlnLysMetAsnHisLeu 80
OY 275 GCAATCATGATGATGAGGAGGATGATATCTTTATGGGTTGCTTGAAGTTCAGGCA 334
Db 81 GlyAspMetHisSerGlnGlnValMetSerLeuMetSerSerLeuArgValProSerGln 100
OY 335 TGGTAGAGCAATGTCATTAATGATGATGATGATGATGATGATGATGATGATGATG 394
Db 101 TrpGlnAlaGlnHisLeuHisLysSerHisProAsnArgLeuLeuProAspSerValAsp 120
OY 395 TCGACAGCAGAGCGCTGTGTTACTGCAAGTCAATATGCAAGCGCTTCTGCTGCTTGG 454
Db 121 TrpArgGlnLysGlnValHisGlnValLysTrpGlnGlnLysSerGlnLysGlnLys 140
OY 455 GTTTTCAGCTGTTGGGCTGCTGAGGACAACTGAACTTAAAGTAAAGAGGAACTG 514
Db 141 AlaPheSerAlaValGlnAlaLeuGlnAlaGlnLeuLysLeuLysTrpGlnLysLeuVal 160
OY 515 TCTTGATGTCACAGCAATGTTGATGATGATGATGATGATGATGATGATGATGATG 574
Db 161 SerLeuSerAlaGlnHisLeuValAspLysSerHisGlnLysTrpGlnLysGlnLys 180
OY 575 AATGCGGCTTCATGACACACTGCTTTCACATATATTATGATTAACACGCGCATTCAT 634
Db 181 AsnGlnLysPheMetHisTrpHisAlaPheGlnTrpHisLeuAspAsnLysGlnLys 200
OY 635 GAAATTTGATGCTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATG 694
Db 201 AspAlaSerTrpProTrpLysAlaMetAspClnLysCysClnTrpAspSerLysTrpArg 220
```

```
US-60-453-135-14611
: Sequence 14611, Application US/60453135
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: IAKOUBOVA, Olga
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001456
: CURRENT APPLICATION NUMBER: US/60/453,135
: NUMBER OF SEQ ID NOS: 2003-03-10
: SOFTWARE: FASTSEQ for Windows version 4.0
: SEQ ID NO 14611
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-453-135-14611

Alignment Scores:
Pred. No.: 1,19e-150 Length: 331
Score: 1571.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.67% Indels: 1
DB: Gaps: 0

US-10-010-580-1 (1-1126) x US-60-453-135-14611 (1-331)

OY 36 ATGAATGGCTAGTGGGCTGCTCCCTGTCCTGCTATGCGGCAAGTACATATA 95
Db 1 MetLysArgLeuValCysValLeuValCysSerSerAlaValAlaGlnLeuHisLys 20
OY 96 GATCCGACATCTGATCATCATCTGAAATCTGTGAAAGAAACCTACAGCAATATC-AAG 154
Db 21 AsprProthLeuAspHisHisIstPrHisLeuTrpHisLysTrpHisLysGlnTrpLys 40
OY 155 GAGACGATGAGGAGGATGACGAGGAGTCTATCTGGGAAAAAATCTMAAATTTGTATG 214
Db 41 GlnLysAsnGlnGlnAlaValAlaArgLeuLeuTrpGlnLysAsnLeuLysPheValMet 60
OY 215 CTTCACAACTTCGCAACACTGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 274
Db 61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTrpAspLeuGlnLysMetAsnHisLeu 80
OY 275 GGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 334
Db 81 GlyAspMetHisSerGlnGlnValMetSerLeuMetSerSerLeuArgValProSerGln 100
```

```

OY 335 TGGCAGAGAAATGTCATTATAGTCGAAACTCTATCAAAATTCCTGATTCCTGGAC 394
DB 101 TTPGlnAArgAsnIleThrTyrIysSerAsnProAsnArgIleLeuProAspSerValAsp 120
OY 395 TGGCAGAGAGAAAGGCTGTACTGTAAGTGAATACGAGGCTTCCTGTCGCTGCTGG 454
DB 121 TTPATGgIulysGlyCysValIthrgIuValIysTyrGlnGlySerCysGlyAlaCysTyr 140
OY 455 GCTTCACGCTGTGGGGCCCTGGACAGCAGCACTGAAGCTAAAGAGGAAAGCTGGTG 514
DB 141 AlaPheSerAlaValGlyAlaLeuGlnIleAsnLeuIysLeuIysThrGlyIysLeuVal 160
OY 515 TCTTCAGATGTCAGAGAACTGTGTAGATTCGTCACATGGAATAATATGGCAATAAAGGCTGC 574
DB 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlnIysTyrGlyAlaAsnIysGlyCys 180
OY 575 AATGGCGCTTCATGACAACTGCTTTCAGATATATATGATTAACAAAGGCAATTCATTCGA 634
DB 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleIleAspAsnIysGlyIleAspSer 200
OY 635 GAAGCTTCCTATCCCTACAAAGCCATGAATGGAAAGTGCAGATATAGTCAAAAAAGGCA 694
DB 201 AspAlaSerTyrProTyrIlysalAmetspGlnIlyCysGlnIlyTyrAspSerIlyTyrArg 220
OY 695 GCTGCCACATGTTCAAGTATACGTGAACCTCCCTTGGCAGTGAAGATGCCCTTAAAGCA 754
DB 221 AlaAlaThrCysSerIlyTyrThrGlnLeuProTyrGlyArgGluAspValLeuIysGlu 240
OY 755 GCTGGGCGCAATAAAGGACCTGTCGTCGCTATAGATGCGGACCACTATTCCTTCTTC 814
DB 241 AlaValAlaAsnIlySglIyProValSerValGlyAlaAspAlaArgHisProSerPhePhe 260
OY 815 CTGACAGAAAGTGTGTCTACTATGAAACCATCCGTGTACTGCAATGTGAATCATGAGATA 874
DB 261 LeuTyrArgSerGlyValIlyTyrGlnProSerCysThrGlnAsnValAlaAsnIlyGlyAla 280
OY 875 TTACGCTGGCTATGCTAACCTTAATGGAAAGCACTAGCTGCTTGTGAAAGACGCTGG 934
DB 281 LeuValValGlyTyrGlyAspLeuAsnGlyIysGlyTyrPheValIlyAsnSerTyr 300
OY 935 GCGCTCAACTTTGGCAGCAAGATATATTCGATGCGCAAGAAACAGTGCAGAAATCACAGT 994
DB 301 GlyHisAsnPheGlyGlnGlyIlyTyrIleArgMetAlaArgAsnIlySglIyAsnHisCys 320
OY 995 GGATGTGCTAGTATGCTTTACCCGAAATG 1027
DB 321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 9
US-60-453-050-14611
: Sequence 14611, Application US/60453050
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: APPLICANT: LUKE, May
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: C0001457
: CURRENT APPLICATION NUMBER: US/60/453,050
: CURRENT FILING DATE: 2003-03-10
: NUMBER OF SEQ ID NOS: 82762
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14611
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-453-050-14611

Alignment Scores:
Pred. No.: 1,19e-150 Length: 331
Score: 1571.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.01% Mismatches: 22

```

```

Query Match: 78.67% Indels: 1
DB: 7 Gaps: 0
US-10-010-580-1 (1-1126) x US-60-453-050-14611 (1-331)

OY 36 ATGAATGGCTAGTGGGCTGCTCCCTGTCGCTGCTATGCAAGTGGCAGCAATGAA 95
DB 1 MetUlsArgLeuValCysValIleuValCysSerSerIleValAlaGlnIleuHisIys 20
OY 96 GATCCACCTGTGATCATCACTGGAATCTGTGNAAGAAACCTACAGCAAAATC-AAG 154
DB 21 AspProThrIleuAspHisIleStrPHisLeuThrPylsYsThrIyrgIyGlySglnIyIys 40
OY 155 CAAGAGATAGAGGAAGTACGACAGCGCTGTATCTGGGAAAAAATCTAAATTTGTGATG 214
DB 41 GlnIysAsnGlnGlnAlaValAlaArgGlyLeuIleIleTropIlyAsnLeuIysPheValMet 60
OY 215 CTTTCACAAATGTCAGCACTCAATGCGAATGCAATCATGATATGATGCAATGCAACATTCG 274
DB 61 LeuHisAsnIleuGlnHisSerMetGlyMetHisSerTyrAspLeuGlnIlyMetAsnHisIleu 80
OY 275 GGAGACATGACTGCTGAAGAAGTGAATCTTTGATGGGTTCCCTGAGAGTTCCACAGCA 334
DB 81 GlyAspMetThrSerGlnGlnIleValMetSerLeuMetSerSerLeuArgValProSerGln 100
OY 335 TGGCAGAGAAATGTCATTATAGTCAAACTCTATACAGAAATTCGCTGATTCGTCGAC 394
DB 101 TTPGlnAArgAsnIleThrTyrIysSerAsnProAsnArgIleLeuProAspSerValAsp 120
OY 395 TGGCAGAGAGAGGCTGTACTGTAAGTGAATACGAGGCTTCCTGTCGCTGCTGG 454
DB 121 TTPATGgIulysGlyCysValIthrgIuValIysTyrGlnGlySerCysGlyAlaCysTyr 140
OY 455 GCTTCACGCTGTGGGGCCCTGGAAAGCACTGAAGTGAATGAAGAGGAGCTGGTG 514
DB 141 AlaPheSerAlaValGlyAlaLeuGlnIleAsnLeuIysLeuIysThrGlyIysLeuVal 160
OY 515 TCTTCAGATGTCAGAGAACTGTGTAGATTCGTCACATGGAATAATATGGCAATAAAGGCTGC 574
DB 161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGlnIysTyrGlnAsnIlySglIyCys 180
OY 575 AATGGCGCTTCATGACAACTGCTTTCAGATATATATGATTAACAAAGCAGCTATTCGA 634
DB 181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleIleAspAsnIysGlyIleAspSer 200
OY 635 GAAGCTTCCTATCCCTACAAAGCCATGAATGGAAGTGCAGATATGATCAAAAGGCA 694
DB 201 AspAlaSerTyrProTyrIlysalAmetspGlnIlyCysGlnIlyTyrAspSerIlyTyrArg 220
OY 695 GCTGCCACATGTTCAAGTATACGTGAACCTCCCTTGGCAGTGAAGATGCCCTTAAAGCA 754
DB 221 AlaAlaThrCysSerIlyTyrThrGlnLeuProTyrGlyArgGluAspValLeuIysGlu 240
OY 755 GCTGGGCGCAATAAAGGACCTGTCGTCGCTATAGATGCGGACCACTATTCCTTCTTC 814
DB 241 AlaValAlaAsnIlySglIyProValSerValGlyAlaAspAlaArgHisProSerPhePhe 260
OY 815 CTGACAGAAAGTGTGTCTACTATGAAACCATCCGTGTACTGCAATGTGAATCATGAGATA 874
DB 261 LeuTyrArgSerGlyValIlyTyrGlnProSerCysThrGlnAsnValAlaAsnIlyGlyAla 280
OY 875 TTACGCTGGCTATGCTAACCTTAATGGAAAGCACTAGCTGCTTGTGAAAGACGCTGG 934
DB 281 LeuValValGlyTyrGlyAspLeuAsnGlyIysGlyTyrPheValIlyAsnSerTyr 300
OY 935 GCGCTCAACTTTGGTGAAGCAAGATATATTCGATGCGCAAGAAACAGTGCAGAAATCACAGT 994
DB 301 GlyHisAsnPheGlyGlnGlyIlyTyrIleArgMetAlaArgAsnIlySglIyAsnHisCys 320
OY 995 GGATGTGCTAGTATGCTTTACCCGAAATG 1027
DB 321 GlyIleAlaSerPheProSerTyrProGluIle 331

```



```

335 TGGCAGAAATGCTACTTATAGTCAAACTCTAATGAAATTCGCTGATTCGTGGAC 394
101 TTPGlnATgAsnIleThrTyrLysSerAsnProAsnArgIleLeuProAspSerValAsp 120
335 TGGACAGAGAAAGGCTGTCTTACTGAAGTGAATACAGAGGCTTCTGTGCTGTGGTGG 454
121 TTPArgGluLysGlyCysValThrGluValLysTyrGlnGlySerCysGlyAlaCysTrp 140
455 GCTTTCACGCGTGTGGGGGCGCTGGAAAGCAGCACTGAAGCTTAAAAACAGCAACTGTC 514
141 AlaPheSerAlaValGlyAlaLeuGluAlaGlnLeuLysLeuLysThrGlyLysLeuVal 160
515 TCTCTGAGTGCACAGAACTTGTAGATTCCTCAACTGAAAAATATGGGAATAAAGGCTGC 574
161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGluLysTyrGlyAsnLysGlyCys 180
575 AATGGCGGTTTCATGACACAGCTTCTCCAGATATATATATATATATATATATATATAT 634
181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleLeuAspAsnLysGlyIleAspSer 200
635 GAAGCTTCCTATCCCTACAAAGCCATGATGAAAGTGCAGATATGACTCAAAAAAGCGA 694
201 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 220
695 GCTGGCCATGTGTAAGTATATACGAACTTCCCTTGGCAGTGAAGTGCCTTAAAGAA 754
221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu 240
755 GCTGTGGCAATAAAGACCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 814
241 AlaValAlaAsnLysGlyLysProValSerValGlyAlaAspAlaArgHisProSerPhePhe 260
815 CTGTACAGAACTGCTGTCTACTATGAACCATCTGTACTGACAAATGTAATCATAGAGTA 874
261 LeuTyrArgSerGlyValIleTyrGluProSerCysThrGlnAsnValAsnHisGlyVal 280
875 TTATGGTGGCTGTGTGTAACCTTAATGGGAAAGACACTGCTTGGAAACAGCTGG 934
281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGluTyrTrpLeuValLysAsnSerTrp 300
935 GGCTCAACTTTGGTGAACCAAGATATATTCGATGCGCAAGAAACAGTGAATCACTGT 994
301 GlyHisAsnPheGlyGlnGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
995 GCGATGCTAGTATTCCTCTTACCAGAAATC 1027
321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 12
US-60-466-412-14611
: Sequence 14611, Application US/60466412
: GENERAL INFORMATION:
: APPLICANT: CARGILL, Michele
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: C1001466
: CURRENT APPLICATION NUMBER: US/60/466,412
: NUMBER OF SEQ. ID NOS: 429241
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ. ID NO 14611
: LENGTH: 331
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-466-412-14611

Alignment Scores:
Pred. No.: 1,19e-150 Length: 331
Score: 1571.00 Matches: 290
Percent Similarity: 93.35% Conservative: 19
Best Local Similarity: 87.61% Mismatches: 22
Query Match: 78.67% Indels: 1
```

```

DB: 7 Gaps: 0
US-10-010-580-1 (1-1126) x US-60-466-412-14611 (1-331)

36 ATGAATGAGTGTAGTGGCGCTGCTGCTGCTATGCAAGTGGCAGCAATACATAA 95
1 MetLysArgLeuValAlaCysValLeuValAlaCysSerSerAlaValAlaGlnLeuHisLys 20
96 GATCCCACTGTGATCATCATCTGAAATCTCTGGAAGAAAACTTACAGCAAAATC -AAG 154
21 AspProThrLeuAsnPhisHisThrPheLysLeuTrpLysLysThrTyrGlyLysGlnTyrLys 40
155 GAAGAGAAATGAGGAAGTACAGCGGCTCTCATCTGGGAAAAAATTTAAATTTGTGATG 214
41 GluLysAsnGluAlaValAlaArgArgLeuIleTrpGluLysAsnLeuLysPheValIleMet 60
215 CTTTCACAATCTGGAACACTCAATGGAATTCATTCATATGATCAGTACGATGCAACATTCG 274
61 LeuHisAsnLeuGlnHisSerMetGlyMetHisSerTyrAspLeuGluMetAsnHisLeu 80
275 GGAAGATGACCTGTGTAAGCAAGTATATCTTTATGAGGTTCCCTGAGAGTTCGCCAACAA 334
81 GlyAspMetThrSerGlnGluValIleMetSerLeuMetSerSerLeuValArgValProSerGln 100
335 TGGCAGAGAAATGTCACTTATAGTCAAACTCTAATCAGAAATTCGCTGATTCGTGAC 394
101 TTPGlnArgAsnIleThrTyrLysSerAsnProAsnArgIleLeuProAspSerValAsp 120
395 TGGACAGACAGAGGCTGTGTACTGTAAGTGAATATACAGAGTTCCTTGTGCTTGTGG 454
121 TTPArgGluLysGlyCysValThrGluValLysTyrGlnGlySerCysGlyAlaCysTrp 140
455 GCTTTCACGCGTGTGGGGGCGCTGGAAGCAACCTGAACCTTAAAAACAGAAAGCTGTCG 514
141 AlaPheSerAlaValAlaGlyAlaLeuGlnAlaGlnLeuLysLeuLysThrGlyLysLeuVal 160
515 TCTCTGAGTGCACAGAACTTGTAGATTCCTCAACTGAAAAATATGGGAATAAAGCTTGC 574
161 SerLeuSerAlaGlnAsnLeuValAspCysSerThrGluLysTyrCysLysLysCys 180
575 AATGGCGGTTTTCATGACACCTGCTTCCAGTATATATATATATATATATATATATATAT 634
181 AsnGlyGlyPheMetThrThrAlaPheGlnTyrIleLeuAspAsnLysGlyIleAspSer 200
635 GAAGCTTCCTATCCCTACAAAGCCATGATGAAAGTGCAGATATGACTCAAAAAAGCGA 694
201 AspAlaSerTyrProTyrLysAlaMetAspGlnLysCysGlnTyrAspSerLysTyrArg 220
695 GCTGGCCATGTGTAAGTATATACGAACTTCCCTTGGCAGTGAAGTGCCTTAAAGAA 754
221 AlaAlaThrCysSerLysTyrThrGluLeuProTyrGlyArgGluAspValLeuLysGlu 240
755 GCTGTGGCAATAAAGACCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 814
241 AlaValAlaAsnLysGlyLysProValSerValGlyAlaAspAlaArgHisProSerPhePhe 260
815 CTGTACAGAACTGCTGTCTACTATGAACCATCTGTACTGACAAATGTAATCATAGAGTA 874
261 LeuTyrArgSerGlyValIleTyrGluProSerCysThrGlnAsnValAsnHisGlyVal 280
875 TTATGGTGGCTGTGTGTAACCTTAATGGGAAAGACACTGCTTGGAAACAGCTGG 934
281 LeuValValGlyTyrGlyAspLeuAsnGlyLysGluTyrTrpLeuValLysAsnSerTrp 300
935 GGCTCAACTTTGGTGAACCAAGATATATTCGATGCGCAAGAAACAGTGAATCACTGT 994
301 GlyHisAsnPheGlyGlnGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
995 GCGATGCTAGTATTCCTCTTACCAGAAATC 1027
321 GlyIleAlaSerPheProSerTyrProGluIle 331

RESULT 13
```



```
Db      ||||||| 261 leuYrAqSerGlyValTyrGluProSerCysThrGlnAsnValAsnHisGlyVal 280
QY      875 TTATGATTGGTTATGTTAACTTAATGGAAAGACTACTGCTTGTGAAAAACAGCTGG 934
Db      281 leuValValGlyTyrGlyAspIleuAsnGlyLysGluTyrTrpIleuValLysAsnSerTrp 300
QY      945 GGCCTCAACTTTGGTGGACCAAGCATATATCGATGGCAAGAAACACTGCATAATCACTGT 994
Db      301 GlnHisAsnPheGlyGluGluGlyTyrIleArgMetAlaArgAsnLysGlyAsnHisCys 320
QY      995 GCGATTGCTTACTTATCCCTCTTACCACGAATC 1027
Db      321 GlyIleAlaSerPheProSerTyrProGluIle 331
```

Search completed: June 27, 2003, 11:50:01
Job time : 141 secs